OPIC
OFFICE DE LA PROPRIÉTÉ INTELLECTUELLE DU CANADA

CIPO
CANADIAN INTELLECTUAL PROPERTY OFFICE

(12) (19) (CA) **Demande-Application**

(21) (A1) **2,252,064**
(22) 1998/11/20
(43) 2000/05/20

(72) BURRELL, Paul Christopher, AU
(72) BLACKALL, Linda Louise, AU
(72) KELLER, Jurg, AU
(71) CRC FOR WASTE MANAGEMENT AND POLLUTION CONTROL LIMITED, AU

(51) Int.Cl.^{6} C12N 1/20, C12Q 1/68, C02F 3/34, C07H 21/04

(54) **MICROORGANISMES OXYDANT LES NITRITES DANS L'EAU**

(54) **AQUATIC NITRITE OXIDISING MICROORGANISMS**

(57) The invention relates to the nitrification of wastewater and identification of microorganisms capable of participating in this process. Specifically, the invention provides a consortium of microorganisms capable of nitrite oxidation in wastewater, which consortium is enriched in members of the Nitrospira phylum. The invention also provides oligonucleotide primers and probes for the amplification or detection of DNA, kits comprising the primers and probes, and methods of detection and quantitating species in a sample.

ABSTRACT

The invention relates to the nitrification of wastewater and identification of microorganisms capable of participating in this process. Specifically, the invention provides a consortium of microorganisms capable of nitrite oxidation in wastewater, which consortium is enriched in members of the *Nitrospira* phylum. The invention also provides oligonucleotide primers and probes for the amplification or detection of *Nitrospira* DNA, kits comprising the primers and probes, and methods of detection and quantitating *Nitrospira* species in a sample.

AQUATIC NITRITE OXIDISING MICROORGANISMS

TECHNICAL FIELD

This invention relates to the removal of nitrogenous compounds from wastewater. In particular, the invention relates to an isolated consortium of microorganisms capable of nitrification of wastewater. The invention also relates to methods of identifying microorganisms capable of nitrification of wastewater and oligonucleotide primers and DNA probes suitable for use in the methods.

INTRODUCTION

The removal of nitrogenous compounds from sewage effluents is an important aspect in the remediation of wastewaters. The presence of ammonia, nitrite and nitrate in wastewater discharges can cause numerous problems ranging from eutrophication (Meganck and Faup, 1988) of the receiving aquatic environment to aspects of public health concern such as nitrate contamination of drinking water. Nitrogen is biologically removed from wastewaters in a two step process of nitrification (ammonia oxidised to nitrate) (Randall, 1992; Robertson and Kuenen, 1991) and denitrification (nitrate reduced to dinitrogen gas that dissipates into the atmosphere) (Blackburn, 1983; Robertson and Kuenen, 1991). Nitrification is the first and most sensitive step of the process and can be further subdivided into two steps: ammonia oxidation to nitrite and nitrite oxidation to nitrate. The two steps are carried out by separate bacterial groups and for both groups, the total diversity of organisms with this phenotype is small.

Therefore, nitrification is a process where reduced nitrogen compounds, generally ammonium (NH_4^+), are microbiologically oxidised to nitrate (NO_3^-) via nitrite (NO_2^-) under aerobic conditions (Halling-Sørensen and Jørgensen, 1993). The overall reactions and possible organisms responsible are:

$$2NH_4^+ + 3O_2 \xrightarrow{\textit{Nitrosomonas}} 2NO_2^- + 2H_2O + 4H^+ + \text{biomass}$$

$$2\ NO_2^- + O_2 \xrightarrow{\textit{Nitrobacter}} 2NO_3^- + \text{biomass}$$

The Gram negative chemoautotrophic nitrite oxidising bacteria are physiologically distinct, as they all possess the ability to use nitrite as their energy source and to assimilate CO_2, via the Calvin Benson cycle, as a carbon source for cell growth (Bock *et al.*, 1992). For each molecule of CO_2 fixed, 100 molecules of nitrite need to be oxidized, emphasising the high energy demands placed on these cells. The overall stoichiometry of nitrite oxidation is (Halling-Sørensen and Jørgensen, 1993):

$$400\ NO_2^- + NH_4^+ + 4H_2CO_3 + HCO_3^- + 195\ O_2 \longrightarrow C_5H_7NO_2 + 3H_2O + 400\ NO_3^-$$

These bacteria can typically also use nitric oxide (NO) instead of NO_2^- as an electron source (Bock *et al.*, 1992). Not all of the known nitrifying bacteria are obligate chemoautotrophs. In fact, many strains of *Nitrobacter* can grow well as heterotrophs, where both energy and carbon are obtained from organic carbon sources, or mixotrophically (a combination of both autotrophic and

heterotrophic behaviour). These bacteria are collectively known as facultative chemoautotrophs. Therefore, bacterial strains can grow three ways; aerobically and autotrophically, aerobically and mixotrophically or anaerobically and heterotrophically. In mixotrophic growth, NO_2^- is oxidized in preference to organic carbon substrates like acetate, pyruvate and glycerol. Both autotrophic and heterotrophic growth is usually slow and inefficient.

As a generalisation, most strains of *Nitrobacter* seem to be able to grow faster as mixotrophs than as heterotrophs and faster heterotrophically or chemo-heterotrophically than chemoautotrophically.

Four genera are currently recognised: *Nitrobacter, Nitrospina, Nitrococcus* and *Nitrospira* (Halling-Sørensen and Jørgensen, 1993). *Nitrospina* and *Nitrococcus* are unable to grow heterotrophically or mixotrophically (Bock *et al.*, 1992). One species of *Nitrospira*, *Nitrospira marina*, can grow autotrophically and mixotrophically, (Bock *et al.*, 1992) whereas *Nitrospira moscoviensis* is an obligate autotroph (Ehrich, *et al.*, 1995). These nitrite oxidizers have also been conventionally classified based on phenotypic characters like their cell shape and the ultrastructure of their intracytoplasmic membranes. Doubling times of *Nitrobacter* can range from 12 to 59 hours, or even as long as 140 hours (Halling-Sørensen and Jørgensen, 1993). These are therefore very slow growing bacteria.

In wastewater treatment systems, *Nitrosomonas* (an ammonia oxidizer) and *Nitrobacter* (a nitrite oxidizer) are the two autotrophs presumed to be responsible for nitrification because they are the commonest ammonia and nitrite oxidizers isolated from these environments (Halling-Sørensen and Jørgensen, 1993). Although ammonia oxidizers have been intensively studied by the use of molecular methods (Wagner *et al.*, 1995; Wagner *et al.*, 1996), the nitrite oxidizers have not been similarly investigated. Since the microorganisms responsible for nitrite oxidation in wastewater treatment plants were presumed to be from the genus *Nitrobacter*, mathematical modeling of the process has used data relevant to this genus. However, fluorescent *in situ* hybridization (FISH) probing of activated sludge mixed liquors with *Nitrobacter* specific probes (Wagner *et al.*, 1996) could not confirm the presence of these organisms suggesting that they were not responsible for this major component of nitrogen remediation. Indeed, *Nitrobacter* could not be found in other aquatic environments (Hovanec and DeLong, 1996) when specific FISH probes were employed. It was speculated that other bacteria were likely responsible for nitrite oxidation (Hovanec and DeLong, 1996; Wagner *et al.*, 1996).

Knowledge of the microorganisms responsible for nitrification of wastewater is desirable for the efficient management of treatment systems. It would also be advantageous to have available biomass which can be added to a system to implement or improve nitrification. However, as indicated above, there is no certainty in the art as to the actual microorganisms responsible for nitrification nor are there methods available for identifying such organisms.

SUMMARY OF THE INVENTION

It is an object of the invention to provide a consortium of microorganisms that can be used for nitrification of wastewater.

A further object of the invention is to provide a method of identifying microorganisms capable of nitrification of wastewater.

According to a first embodiment of the invention, there is provided a consortium of microorganisms capable of nitrite oxidation in wastewater, which consortium is enriched in members of the *Nitrospira* phylum.

According to a second embodiment of the invention, there is provided an oligonucleotide primer for PCR amplification of *Nitrospira* DNA, said primer comprising at least 12 nucleotides having a sequence selected from:

(i) any one of SEQ ID NO: 1 to SEQ ID NO: 13; or

(ii) a DNA sequence having at least 92% identity with any one of SEQ ID NO: 1 to SEQ ID NO: 13.

According to a third embodiment of the invention, there is provided a primer pair for PCR amplification of *Nitrospira* DNA, said primer pair comprising:

(a) a first oligonucleotide of at least 12 nucleotides having a sequence selected from one strand of a bacterial 16S rDNA gene; and

(b) a second oligonucleotide of at least 12 nucleotides having a sequence selected from the other strand of said 16S rDNA gene downstream of said first oligonucleotide sequence; wherein at least one of said first and second oligonucleotides is selected from:

(i) any one of SEQ ID NO: 1 to SEQ ID NO: 13; or

(ii) a DNA sequence having at least 92% identity with any one SEQ ID NO: 1 to SEQ ID NO: 13.

According to a fourth embodiment of the invention, there is provided a probe for detecting *Nitrospira* DNA, said probe comprising at least 12 nucleotides having a sequence selected from:

(i) any one of SEQ ID NO: 1 to SEQ ID NO: 13; or

(ii) a DNA sequence having at least 92% identity with any one of SEQ ID NO: 1 to SEQ ID NO: 13.

According to a fifth embodiment of the invention, there is provided a kit comprising:

at least one primer according to the second embodiment;

at least one primer pair according to the third embodiment; or

at least one probe according to the fourth embodiment.

According to a sixth embodiment of the invention, there is provided a method of detecting a *Nitrospira* species in a sample, said method comprising the steps of:

(a) lysing cells in said sample to release genomic DNA;

(b) contacting denatured genomic DNA from step (a) with a primer pair according to the third embodiment;

(c) amplifying *Nitrospira* DNA by cyclically reacting said primer pair with said DNA to produce an amplification product; and

(d) detecting said amplification product.

According to a seventh embodiment of the invention, there is provided a method of quantitating the level of a *Nitrospira* species in a sample, said method comprising the steps of:

(a) lysing cells in said sample to release genomic DNA;

(b) contacting denatured genomic DNA from step (a) with a primer pair according to the third embodiment;

(c) amplifying *Nitrospira* DNA by cyclically reacting said primer pair with said DNA to produce an amplification product; and

(d) detecting said amplification product and quantitating the level of said product by comparison with at least one reference standard.

According to an eighth embodiment of the invention, there is provided a method of detecting a *Nitrospira* species in a sample, said method comprising the steps of:

(a) lysing cells in said sample to release genomic DNA;

(b) contacting denatured genomic DNA from step (a) with a labeled probe according to the fourth embodiment under conditions which allow hybridisation of said genomic DNA said probe;

(c) separating hybridised labeled probe and genomic DNA from unhybridised labeled probe; and

(d) detecting said labeled probe-genomic DNA hybrid.

According to a ninth embodiment of the invention, there is provided a method of detecting cells of a *Nitrospira* species in a sample, said method comprising the steps of:

(a) treating cells in said sample to fix cellular contents;

(b) contacting said fixed cells from step (a) with a labeled probe according to the fourth embodiment under conditions which allow said probe to hybridise with RNA within said fixed cell;

(c) removing unhybridised probe from said fixed cells; and

(d) detecting said labeled probe-RNA hybrid.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a graph showing influent and effluent NO_2-N concentrations for an automated laboratory-scale reactor operating as a sequencing batch reactor at 2 cycles/day with strong selection for nitrite oxidising biomass (NOSBR).

Figure 2 is a graph showing influent and effluent NO_2-N concentrations of the NOSBR operating at 4 cycles/day.

Figure 3 is a graph of mixed liquor nitrite-N concentrations during the react period of the NOSBR cycle for attached growth and for suspended growth.

Figure 4 is a graph showing nitrite-N and nitrate-N concentrations in the mixed liquor during the react period of the NOSBR.

Figure 5 ia a graph showing mixed liquor nitrite-N concentrations during the react period in three stages of the NOSBR operated at 2 cycles/day with different concentrations of nitrite in the feed.

Figure 6 is a graph of mixed liquor nitrite-N concentrations during the react period in three representative cycles during operation of the NOSBR at 4 cycles/day.

Figure 7 is an evolutionary distance tree derived from a comparison of 16S rDNA sequences from nitrite oxidising bacteria and clone sequences from three different 16S rDNA clone libraries (RC, GC, and SBR).

Figure 8 is an alignment of sequences of 16S rDNA from *Nitrospira* clones identified in a nitrite-oxidising SBR and from other sources.

Figure 9 depicts the results of agarose gel electrophoresis of PCR-amplified DNA using genomic DNA from various *Nitrospira* clones as template.

BEST MODE AND OTHER MODES OF CARRYING OUT THE INVENTION

The following abbreviations are used hereafter:

SBR	sequencing batch reactor
NOSBR	nitrite oxidising SBR
NOM	nitrite oxidising medium
HRT	hydraulic retention time
MLSS	mixed liquor suspended solids
BNR	biological nutrient removal
DO	dissolved oxygen
PCR	polymerase chain reaction
REA	restriction enzyme analysis
OTU	operational taxonomic unit
bp(s)	base pair(s)

The one-letter code for nucleotides in DNA conforms to the IUPAC-IUB standard described in *Biochemical Journal* 219, 345-373 (1984).

The term “comprise”, or variations of the term such as “comprises” or “comprising”, are used herein to denote the inclusion of a stated integer or stated integers but not to exclude any other

integer or any other integers, unless in the context or usage an exclusive interpretation of the terms is required.

The present inventors have developed a specific nitrifying biomass that is largely comprised of bacteria that are most closely related to *Nitrospira moscoviensis*. It is believed that a range of species of *Nitrospira* are involved in the process. The inventors have shown that these bacteria are likely to be more dominant in reactors with good nitrification performance than bacteria from the genus *Nitrobacter*. A range of studies have failed to find *Nitrobacter* in nitrifying processes (Hovanec & DeLong, 1996; Wagner *et al.*, 1996) and evidence is provided below that the organisms responsible for this important biochemical reaction in wastewater treatment processes (both suspended and attached growth processes) are from the *Nitrospira* phylum in the domain *Bacteria*.

With reference to the first embodiment of the invention, the nitrifying biomass can be produced by presenting a feed comprising nitrite, dissolved oxygen and dissolved carbon dioxide but which is free of organic carbon to seed sludge from any sewage plant exhibiting nitrification. The seed sludge is advantageously from a domestic wastewater treatment plant but can also be from an abattoir wastewater treatment plant. The nitrite component of the feed can be as low as about 400 mg/L nitrite-N. The oxygen and carbon dioxide can conveniently be provided as air bubbled through the solution.

Turning to the second embodiment of the invention, oligonucleotide primers typically have a length of about 12 to 50 nucleotides. A preferred length is 12 to 22 nucleotides. Particularly preferred primers are the following:

5' CGGGAGGGAAGATGGAGC 3'	(SEQ ID NO: 14)
5' CCAACCCGGAAAGCGCAGAG 3'	(SEQ ID NO: 15)
5' AGCCTGGCAGTACCCTCT 3'	(SEQ ID NO: 16)

Oligonucleotide primer pairs according to the third embodiment of the invention comprise an oligonucleotide primer that will anneal to one strand of the target sequence and a second oligonucleotide primer which will anneal to the other, complementary, strand of the target sequence. It will be appreciated that the second oligonucleotide primer must anneal to the complementary strand downstream of the first oligonucleotide primer sequence, which occurs in the complementary strand, to yield a double stranded amplification product in the PCR. The amplification product is of a size that facilitates detection. Typically, the first and second oligonucleotide primer sites in the target DNA are separated by 50 to 1,400 bps. A preferred separation is 400 to 1,000 bps.

The probes of the fourth embodiment, as indicated above, can have a size as small as 12 nucleotides. Typically, however, probes have a length of 15 to 50 nucleotides. A preferred probe length is 15 to 22 nucleotides, particularly for *in situ* hybridisation according to the method of the ninth embodiment.

The oligonucleotide primers included in kits according to the fifth embodiment of the invention can be individual oligonucleotide primers appropriate for the detection of *Nitrospira* or a primer pair. Oligonucleotide primer pairs are advantageously provided as compositions. Additional oligonucleotide primers can also be included in kits for use in control reactions. For detection purposes, DNA probes can also be included in kits.

Kits according to the fifth embodiment of the invention can further comprise reagents used in PCR and hybridisation reactions. Such reagents include buffers, salts, detergents, nucleotides and thermostable polymerase. Such reagents are advantageously provided as solutions to facilitate execution of PCR or hybridisation. Solutions can be compositions comprising a number of reagents as is well known in the art.

The general techniques used in the methods of the sixth to ninth embodiments, and factors to be considered in selecting PCR primers and probes, will be known to those of skill in the art. Such techniques are described, for example, in Sambrook *et al.* (1989) and Stackebrandt and Goodfellow (1991), the entire contents of which are incorporated herein by cross reference. Particularly relevant chapters in Stackebrandt and Goodfellow are Chapter 7, "The Polymerase Chain Reaction" by S. Giovannoni, and Chapter 8, "Development and Application of Nucleic Acid Probes" by D. A. Stohl and R. Amann.

Non-limiting examples of the invention will now be provided.

General Methods

The total community DNAs from the NOSBR sludge (RC) and the seed sludge (GC) were isolated, the 16S rDNAs were polymerase chain reaction (PCR) amplified and cloned using previously published methods (Blackall, 1994; Blackall *et al.*, 1994; Bond *et al.*, 1995). Inserts from 102 clones in the RC library were amplified and grouped by *Hae*III restriction enzyme digestion banding profiles (REA) into operational taxonomic units (OTUs) (Weidner *et al.*, 1996). Clone inserts from representatives of RC OTUs and all 77 clones from the GC library were PCR amplified and partially sequenced (Blackall, 1994) using 530f (Lane, 1991) primer. Inserts from a selection of clones were fully sequenced (Blackall, 1994). Sequence data were analysed according to previously published methods (Blackall *et al.*, 1994) which included BLAST (Altschul *et al.*, 1990) comparisons and phylogenetic analyses (Felsenstein, 1993).

Example 1

Selection of a Nitrifying Biomass

In this example, we describe the use of a laboratory-scale reactor as a sequencing batch reactor (SBR) with strong selection for a nitrite oxidising biomass. Seed sludge was from the Merrimac domestic wastewater treatment plant operated by the Gold Coast City Council and located

at Merrimac, Queensland 4226, Australia. The reactor set-up will be hereafter referred to as the "Nitrite Oxidising SBR", or "NOSBR".

Reactor. A laboratory chemostat with a working volume of 1 L was operated in the dark at 24°C as the NOSBR. The influent nitrite oxidising medium (NOM) was a synthetic waste water mix comprising per L: 400 to 3,200 mg KNO_2, 3.75 g $MgSO_4.7H_2O$, 250 mg $CaCl_2.2H_2O$, 10 g KH_2PO4, 10 g K_2HPO_4, 200 mg $FeSO_4.7H_2O$, and 20 g $NaHCO_3$. The pH of the medium was adjusted to 7.0, but the reactor was not equipped with pH control. Dissolved oxygen was maintained at 1.6-2.0 mg/L and CO_2 was introduced by bubbling air through the liquid in the NOSBR. Surface biomass growth was precluded by regular scrubbing of all solid surfaces with a brush. Four cycles per day giving a hydraulic retention time (HRT) of 12 hr were performed with the following sequences:-

1) Feed of 500 ml of fresh medium - 30 min (0 to 0.5 hr)
2) React (aeration) - 4.5 hr (0.5 to 5 hr)
3) Settle - 40 min (5 to 5.7 hr)
4) Decant 500 ml of supernatant - 20 min (5.7 to 6 hr)
5) Total time per cycle - 6 hr.

Automatic timers controlled the magnetic stirrer (100 rpm), peristaltic pumps (feed and decant), and air pump for the cycles. Sludge biomass was not wasted from the reactor, but periodically, biomass was collected for testing which facilitated maintenance of a relatively steady amount of biomass in the SBR.

At start up, 1 L of mixed liquor suspended solids (MLSS) from a full scale Biological Nutrient Removal (BNR, nitrogen and phosphorus removal) plant was added to the NOSBR which was operated manually with the NOM. Initial manual and then automatic operation with 2-cycles per day (feed - [500 ml] 40 min; react - 10 hr; settle - 40 min; and decant [500 ml] - 40 min) occurred for some months before initiation of the 4-cycles per day scheme (see above).

Monitoring. Chemical analyses of feed, mixed liquor and effluent were regularly done for nitrite-N (NO_2-N), nitrate-N (NO_3-N), and ammonium-N (NH_4^+-N) using spectrometric assays (Merck, Melbourne, Australia). To preclude the removal of excessive biomass, these analyses were done with 2 ml samples. The MLSS of the NOSBR was determined in duplicate 10 ml samples of mixed liquor. These were filtered onto pre-dried Whatman GF/C filters, and then dried to a constant weight at 105 degree C. A pH meter was used to periodically monitor pH in the mixed liquor and effluent. A portable dissolved oxygen (DO) meter and probe were used to periodically monitor the DO in the NOSBR.

Results of operation. Varying influent nitrite levels were employed to study a range of features of the selected nitrite oxidising biomass. The operating data for the influent and effluent nitrite levels

of the NOSBR during the automated 2 cycles/day period are presented in Figure 1 and for the automated 4 cycles/day in Figure 2. The data presented in these figures show that the microbial community are able to remove all the nitrite from the influent in a matter of hours.

Attributes of the NOSBR mixed liquor

1. Suspended versus attached growth - 2 cycles/day. To generate attached growth, the regular scrubbing regime of the reactor was suspended for two weeks. The vast bulk of the biomass was then attached to surfaces in the reactor. The little remaining suspended biomass was discharged from the reactor which was then filled with 1 L of half strength NOM. Regular sampling and nitrite analyses were done during the react period of one cycle with all the biomass attached to the reactor surfaces. The results of this experiment are presented in Figure 3. The results show that suspended biomass has twice the nitrite oxidation rate than the attached biomass but both systems are effective in removing nitrite from the influent.

Following the experiment described in the previous paragraph, the biomass was completely scrubbed from the surfaces to the liquid. The reactor was operated for two cycles with biomass scrubbing. A similar one-cycle study was performed as with the attached growth but with all biomass suspended. The biofilm growth exhibited a nitrite oxidation rate of 29 mg NO_2-N/hr and the suspended growth form showed a rate of 58 mg NO_2-N/hr. It was assumed that the biomass concentration was the same for both studies since none had been removed between them.

2. pH correlation with nitrification. It was observed that when the pH of the effluent fell below 7.4, nitrite-N was present in the effluent. If the pH rose above 7.4 for short periods, no effect to nitrification was observed. Therefore, pH values below 7.4 were detrimental to nitrification.

3. Cyclic studies. Figure 4 shows the results for periodic measurements of nitrite-N and nitrate-N during the react period of the reactor during 2 cycles/day The results presented in these figures show that the bacterial population in the reactor oxidised nitrite to nitrate in a stoichiometric manner with 160 mg/l of nitrite-N being oxidised to 160 mg/l of nitrate-N (170 mg/l at the start of the react period and 330 mg/l when the nitrite-N was exhausted). The rate of nitrite oxidation and nitrate production also appeared to be linear, showing that the oxidation process was not limited by any external factors.

Studies measuring nitrite reaction in the reactor are shown for both 2 cycles/day (Figure 5) and 4 cycles/day operation (Figure 6). The significance of these results is that the biomass is robust in its capacity to oxidise nitrite under a range of operating conditions.

Example 2

The Microbiology of the NOSBR

In this example, we describe the microbiological characterisation of the nitrifying microorganisms present in the biomass selected in the NOSBR described in Example 1. Methods used

in the characterisation have been described by Blackall (1994) and Bond *et al.* (1995), the entire contents of which disclosures are incorporated herein by cross-reference.

Total microbial community DNA from both the seed BNR sludge (GC) and from the reactor after six months of operation (RC) was obtained. The 16S rDNA from each DNA extract were separately amplified by polymerase chain reaction (PCR), and then for each, clone libraries were prepared (Blackall, 1994; Bond *et al.*, 1995).

Inserts from a total of 77 clones from the GC clone library were partially sequenced with the primer 530f and phylogenetically analysed (Blackall *et al.*, 1994) (Table 1). The majority of the clone sequences grouped with the proteobacterial phylum, while 4% (3 clones; GC3, GC86 and GC109) grouped with the phylum *Nitrospira*.

Table 1

Phyla from the Domain Bacteria Represented in the GC Clone Library

Phylum in Domain Bacteria	Percentage in clone library
Proteobacteria	
Alpha	5
Beta	29
gamma	18
delta	4
High mol%G+C Gram positives	10
Low mol%G+C Gram positives	7
Flexibacter/Cytophaga/Bacteroides	5
Nitrospira	4
Planctomycetales	9
Unaffiliated	9

Restriction Enzyme Analysis (REA) of the RC library was done to group clones into operational taxonomic units (OTUs) in advance of partial or complete clone insert sequencing (Weidner *et al.*, 1996). Thirteen different OTUs were found when *Hae*III was employed as the restriction enzyme to digest the inserts from 102 clones. The large majority of the clone inserts (88% or 90 clones) were found in one OTU while the remaining 12% (12 clones) comprised individuals in 12 other OTUs. Each of the clone inserts from the latter 12 OTUs and six of the large former group (RC7, RC11, RC16, RC25, RC73, and RC99) were partially sequenced and phylogenetically analysed. These six and one of the other OTUs (RC90) were found to have partial insert sequences that phylogenetically grouped with the *Nitrospira* phylum. From this analysis, it was concluded that 91 clones or 89% of the clone library originated from bacteria in the *Nitrospira* phylum. In the

phylogenetic analysis, one of the other OTUs (RC44) grouped with *Nitrobacter*. It was concluded that the organisms responsible for nitrification in the NOSBR were likely to be from the *Nitrospira* phylum.

Near complete insert sequence analyses were done for the following clones:

- six RC clones of the original partial sequences - RC7, RC11, RC25, RC73, RC90, and RC99 (RC16 omitted);
- two RC clones from the *Nitrospira* OTU (RC14 and RC19);
- one of the three GC *Nitrospira* clones (GC86); and
- four clones from a clone library prepared by Bond *et al.* (1995) that phylogenetically grouped in the *Nitrospira* phylum.

The data were phylogenetically analysed as shown in Figure 7. The two clone clades would likely comprise two separate species with the RC clones possibly comprising more than one species.

Sequences of clones from the two *Nitrospira* clades were subjected to direct pairwise sequence comparison. The results of this comparison are presented in Table 2. The table is a similarity matrix showing the percent similarity between 16S rDNA sequences of *Nitrospira moscoviensis*, *Nitrospira marina* and 13 near complete sequences from clone inserts from a full scale biological nutrient removal activated sludge plant (GC86), from the NOSBR (RC clone numbers) and from clones for which the partial sequences had been previously reported (SBR clones; Bond *et al.*, 1995). The similarity matrix showed that the first clade (SBR1015, SBR1024, SBR2046, GC86) had an average 16S rDNA comparison value of 99.4% while for the second clade (RC7, RC11, RC14, RC19, RC25, RC73, RC90, RC99, SBR2016), this value was 98.7%. The highest comparative value between an RC clone sequence and *N. moscoviensis* was 93.4% for RC25. From the sequence data analysis, the two clone clades would likely comprise two separate species, with the RC clones possibly comprising more than one species.

Sequence data for the SBR, GC and RC clones are presented in Figure 8. In this figure, sequences are divided into blocks with numbers given in square brackets above each block. The clone identification is given at the left of a line of sequence in each block. Dashes represent unknown nucleotides while full stops represent alignment breaks.

The sequences of clones are also presented as sequence listings as follows:

Clone	Sequence Listing Number
SBR1024	1
SBR1015	2
GC86	3
SBR2046	4
RC25	5
RC19	6
SBR2016	7
RC7	8
RC14	9
RC99	10
RC11	11
RC73	12
RC90	13

Table 2

Species or clone	Percent sequence similarity with species of strain number														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. *Nitrospira moscoviensis*															
2. SBR1024	96.3														
3. SBR1015	96.1	**99.6**													
4. GC86	96.1	**99.6**	**99.4**												
5. SBR2046	95.8	**99.3**	**99.4**	**99.2**											
6. RC25	93.4	93.4	93.6	93.6	93.1										
7. RC19	93.2	93.1	93.0	93.2	92.7	**98.8**									
8. SBR2016	93.0	92.7	92.8	92.6	92.4	**99.1**	**98.7**								
9. RC7	92.9	93.1	93.2	92.9	92.8	**98.7**	**98.7**	**98.5**							
10 RC14	92.8	93.0	93.1	93.1	92.7	**98.7**	**98.9**	**98.5**	**99.3**						
11 RC99	92.7	92.9	93.0	93.0	92.6	**98.5**	**98.7**	**98.4**	**99.2**	**99.6**					
12 RC11	92.6	92.8	93.0	92.9	92.5	**98.5**	**98.7**	**98.4**	**99.0**	**99.5**	**99.7**				
13 RC73	92.2	92.5	92.6	92.6	92.1	**98.0**	**98.2**	**97.9**	**98.7**	**99.1**	**99.4**	**99.4**			
14 RC90	92.1	92.1	92.3	92.2	91.8	**98.1**	**98.6**	**98.0**	**98.1**	**98.6**	**98.8**	**98.8**	**99.0**		
15 *Nitrospira marina*	88.7	88.2	88.3	88.3	87.8	88.1	87.6	87.2	87.2	87.1	87.1	87.1	86.5	86.6	
16 *Nitrospira marina*	88.0	88.0	88.2	88.1	87.7	87.9	87.5	87.2	87.2	87.1	87.1	87.1	86.5	86.6	99.9

Example 3

Identification of *Nitrospira* Species

Primers for use in a diagnostic PCR for the *Nitrospira moscoviensis* clade of Figure 7 (see Example 2) were designed from aligned sequence datasets (see Tables 3-5 below).

Table 3 is an alignment of 16S rDNA sequences of *Nitrospira* phylum members and nitrite oxidisers from other bacterial phyla which was used to design the primer MOS457f (SEQ ID NO: 14) for the *Nitrospira mascoviensis* clade. In the table, mismatches with the primer sequence are in bold type and are underlined. The melting temperature calculated for MOS457f was 60°C and a fragment size of approximately 1052 nucleotides was calculated in a PCR with primer 1492r. The MOS457f sequence corresponds to the sequence at positions 440 to 457 of the *E. coli* 16S rDNA gene.

Table 3

Source of Sequence and Number of Sequence in Sequence Listings	Sequence	Mismatches
MOS457f primer (SEQ ID NO: 14)	CGGGAGGGAAGATGGAGC	-
Nitrococcus mobilis (SEQ ID NO: 17)	C**A**G**CC**GGGA**G**GA**AAAGCA**	10
Magnetobacterium bavaricum (SEQ ID NO: 18)	**T**G**TAG**GG**A**AAGATG**AT**G**A**	8
Nitrobacter hamburgensis (SEQ ID NO: 19)	**T**G**T**G**C**GGGAAGAT**AAT**G**A**	7
Nitrospina gracilis (SEQ ID NO: 20)	CGGG**T**GGGAAGA**ACA**A**AA**	6
Nitrospira marina (SEQ ID NO: 21)	C**AT**GAGG**A**AAGAT**AA**AG**T**	6
SBR1015 (SEQ ID NO: 22)	CGG**C**AGGGAAGATGGA**A**C	2
SBR1024 (SEQ ID NO: 22)	CGG**C**AGGGAAGATGGA**A**C	2
SBR2016 (SEQ ID NO: 23)	CGGGAGGGAAGATGGAGC	0
SBR2046 (SEQ ID NO: 24)	C**C**G**C**AGGGAAGATGGA**A**C	3
RC7 (SEQ ID NO: 23)	CGGGAGGGAAGATGGAGC	0
RC11 (SEQ ID NO: 23)	CGGGAGGGAAGATGGAGC	0
RC14 (SEQ ID NO: 23)	CGGGAGGGAAGATGGAGC	0
RC19 (SEQ ID NO: 23)	CGGGAGGGAAGATGGAGC	0
RC25 (SEQ ID NO: 23)	CGGGAGGGAAGATGGAGC	0
RC73 (SEQ ID NO: 25)	CGGGAGGGAAGATGGA**A**C	1
RC90 (SEQ ID NO: 25)	CGGGAGGGAAGATGGA**A**C	1
RC99 (SEQ ID NO: 23)	CGGGAGGGAAGATGGAGC	0
RC44 (*Nitrobacter* clone) (SEQ ID NO: 26)	CG**T**G**C**GGGAAGAT**AAT**G**A**	6
GC86 (SEQ ID NO: 27)	CGG**C**AGGGAAGATGGA**A**C	2
Nitrospira moscoviensis (SEQ ID NO: 28)	CGGGAGGGAAGATGGA**CG**	2

Like Table 3, Table 4 is an alignment of 16S rDNA sequences of *Nitrospira* phylum members and nitrite oxidisers from other bacterial phyla which was used to design the primer MOS638f (SEQ ID NO: 15) for the *Nitrospira moscoviensis* clade. Again, mismatches with the primer sequence are in bold and are underlined. The calculated melting temperature for this primer was 66°C and a fragment size of approximately 873 nucleotides was calculated in a PCR with primer 1492r. The MOS638f sequence corresponds to the sequence at positions 619 to 638 of the *E. coli* 16S rDNA gene.

Table 4

Source of Sequence and Number of Sequence in Sequence Listings	Sequence	Mismatches
MOS638f primer (SEQ ID NO: 15)	CCAACCCGGAAAGCGCAGAG	-
Nitrococcus mobilis (SEQ ID NO: 29)	**T**CAACC**T**GG**G**AA**TT**GCA**TCC**	8
Magnetobacterium bavaricum (SEQ ID NO: 30)	**T**CAACCCGG**G**AA**TT**GC**CTT**G	7
Nitrobacter hamburgensis (SEQ ID NO: 31)	**T**CAAC**T**C**CAG**AA**CT**GC**CTTT**	11
Nitrospina gracilis (SEQ ID NO: 32)	**T**CAACC**GT**G**G**AA**TT**GC**GTTT**	10
Nitrospira marina (SEQ ID NO: 33)	**TT**AACC**G**GGAAAG**GT**C**GAGA**	9
SBR1015 (SEQ ID NO: 34)	C**T**AACCCGGAAAG**T**GC**G**GAG	3
SBR1024 (SEQ ID NO: 34)	C**T**AACCCGGAAAG**T**GC**G**GAG	3
SBR2016 (SEQ ID NO: 35)	CCAACCCG**A**AAAGCGCAGAG	1
SBR2046 (SEQ ID NO: 34)	C**T**AACCCGGAAAG**T**GC**G**GAG	3
RC7 (SEQ ID NO: 36)	CCAACCCGGAAAGCGCAGAG	0
RC11 (SEQ ID NO: 36)	CCAACCCGGAAAGCGCAGAG	0
RC14 (SEQ ID NO: 36)	CCAACCCGGAAAGCGCAGAG	0
RC19 (SEQ ID NO: 36)	CCAACCCGGAAAGCGCAGAG	0
RC25 (SEQ ID NO: 36)	CCAACCCGGAAAGCGCAGAG	0
RC73 (SEQ ID NO: 36)	CCAACCCGGAAAGCGCAGAG	0
RC90 (SEQ ID NO: 36)	CCAACCCGGAAAGCGCAGAG	0
RC99 (SEQ ID NO: 36)	CCAACCCGGAAAGCGCAGAG	0
RC44 (*Nitrobacter* clone) (SEQ ID NO: 37)	**T**CAAC**T**C**CAG**AA**CT**GC**CTTT**	11
GC86 (SEQ ID NO: 34)	C**T**AACCCGGAAAG**T**GC**G**GAG	3
Nitrospira moscoviensis (SEQ ID NO: 38)	CCAACCCGGAAAGCGCAGAG	0

Table 5, is again an alignment of 16S rDNA sequences of *Nitrospira* phylum members and nitrite oxidisers from other bacterial phyla which was used to design the primer MOS635r (SEQ ID

NO: 16) for the *Nitrospira moscoviensis* clade. The melting temperature calculated for this primer was 58°C and a fragment size of approximately 625 nucleotides was calculated in a PCR with primer 27f. The MOS635r sequence corresponds to the sequence at positions 635 to 652 of the *E. coli* 16S rDNA sequence.

Table 5

Source of Sequence and Number of Sequence in Sequence Listings	Sequence	Mismatches
MOS635r primer (SEQ ID NO: 16)	AGCCTGGCAGTACCCTCT	-
Nitrococcus mobilis (SEQ ID NO: 39)	AGCCAAACAGTATCGGAT	7
Magnetobacterium bavaricum (SEQ ID NO: 40)	AGTTAAACAGTTTTCAAG	11
Nitrobacter hamburgensis (SEQ ID NO: 41)	AGACCTTCAGTATCAAAG	9
Nitrospina gracilis (SEQ ID NO: 42)	AGCCGAATAGTTTCAAAC	10
Nitrospira marina (SEQ ID NO: 43)	AGCTGAATAGTTCCTCTC	10
SBR1015 (SEQ ID NO: 44)	AGCCGAGCAGTCCCCTCC	4
SBR1024 (SEQ ID NO: 44)	AGCCGAGCAGTCCCCTCC	4
SBR2016 (SEQ ID NO: 45)	AGCCTGGCAGTACCCTCT	0
SBR2046 (SEQ ID NO: 44)	AGCCGAGCAGTCCCCTCC	4
RC7 (SEQ ID NO: 46)	AGCCTGGCAGTACCCCCT	1
RC11 (SEQ ID NO: 45)	AGCCTGGCAGTACCCTCT	0
RC14 (SEQ ID NO: 45)	AGCCTGGCAGTACCCTCT	0
RC19 (SEQ ID NO: 45)	AGCCTGGCAGTACCCTCT	0
RC25 (SEQ ID NO: 47)	AGCCTGGCAGTACCGTCT	1
RC73 (SEQ ID NO: 45)	AGCCTGGCAGTACCCTCT	0
RC90 (SEQ ID NO: 45)	AGCCTGGCAGTACCCTCT	0
RC99 (SEQ ID NO: 45)	AGCCTGGCAGTACCCTCT	0
RC44 (*Nitrobacter* clone) (SEQ ID NO: 48)	AGATCCTCAGTATCAAAG	10
GC86 (SEQ ID NO: 44)	AGCCGAGCAGTCCCCTCC	4
Nitrospira moscoviensis (SEQ ID NO: 49)	AGCCTGGCAGTACCCTCT	0

The three primers defined above in Tables 3 to 5 were included in separate primer pairs which pairs were then tested in PCR amplifications using genomic DNA from various *Nitrospira* clones as template. The PCRs were carried out according to methods detailed in Sambrook *et al.* (1989) at an annealing temperature of 62°C.

The results of electrophoretic analysis of PCRs on an agarose gel are presented in Figure 9. Details of the material analysed in each lane of the gel are given in Table 6. The marker DNA was

*Hae*III-digested ϕX174 DNA. The sizes of the ϕX174 fragments are given on the left-hand side of the figure.

Table 6

Lane	Primer pair used	Mismatches between primer and template
1	(*Hae*III-digested ϕX174 DNA)	
2	MOS457f, 1492r	0 mismatches with MOS457f
3	MOS457f, 1492r	1 mismatch with MOS457f
4	MOS457f, 1492r	2 mismatches with MOS457f
5	(*Hae*III-digested ϕX174 DNA)	
6	MOS638f, 1492r	0 mismatches with MOS638f
7	MOS638f, 1492r	1 mismatch with MOS638f
8	MOS638f, 1492r	3 mismatches with MOS638f
9	(*Hae*III-digested ϕX174 DNA)	
10	MOS635r, 27f	0 mismatches with MOS635r
11	MOS635r, 27f	1 mismatch with MOS635r
12	MOS635r, 27f	4 mismatches with MOS635r

The results presented in Figure 9 show that an amplicon of the appropriate size was obtained in reactions where there was up to one mismatch between a primer and the template but that no amplicon was produced where there was a greater degree of mismatch.

When the three primer pairs used for the results presented in Figure 9 were used with clone RC44 (closest match to *Nitrobacter*), no amplicons were produced.

The primer NIT3 (Wagner *et al.* 1996; SEQ ID NO: 50) was used in a diagnostic PCR for *Nitrobacter*. NIT3 was designed originally for fluorescent *in situ* hybridisation experiments. The specificity of this primer can be appreciated from the sequence alignment presented in Table 7 which is an alignment of 16S rDNA sequences of *Nitrospira* phylum members and nitrite oxidisers from other bacterial phyla against NIT3. A melting temperature of 60°C was calculated for NIT3 and a fragment size of approximately 1020 nucleotides in a PCR with primer 27f as experimentally determined. The NIT3 sequence corresponds to the sequence at positions 1031 to 1048 of the *E.coli* 16S rDNA gene.

Table 7

Source of Sequence and Number of Sequence in Sequence Listings	Sequence	Mismatches
NIT3 primer (SEQ ID NO: 50)	CCTGTGCTCCATGCTCCG	-
Nitrobacter hamburgensis (SEQ ID NO: 51)	CCTGTGCTCCATGCTCCG	0
Nitrospina gracilis (SEQ ID NO: 52)	CCTGTGCAAGGGCCCCGA	9
Nitrococcus mobilis (SEQ ID NO: 53)	CCTGTCATCCGGTTCCCG	7
Nitrospira moscoviensis (SEQ ID NO: 54)	CCTGAGCACGCTGGTATT	8
Nitrospira marina (SEQ ID NO: 55)	CCTGAGCTCGCTCCCCTT	7
Magnetobacterium bavaricum (SEQ ID NO: 56)	CCTGTGCAAGCTCTCCCT	8
SBR1015 (SEQ ID NO: 57)	CCTGAGCAGGATGGTATT	8
SBR1024 (SEQ ID NO: 57)	CCTGAGCAGGATGGTATT	8
SBR2016 (SEQ ID NO: 58)	CCTGAGCACGCTGGTATT	8
SBR2046 (SEQ ID NO: 57)	CCTGAGCAGGATGGTATT	8
RC7 (SEQ ID NO: 58)	CCTGAGCACGCTGGTATT	8
RC11 (SEQ ID NO: 58)	CCTGAGCACGCTGGTATT	8
RC14 (SEQ ID NO: 58)	CCTGAGCACGCTGGTATT	8
RC19 (SEQ ID NO: 58)	CCTGAGCACGCTGGTATT	8
RC25 (SEQ ID NO: 58)	CCTGAGCACGCTGGTATT	8
RC73 (SEQ ID NO: 58)	CCTGAGCACGCTGGTATT	8
RC90 (SEQ ID NO: 58)	CCTGAGCACGCTGGTATT	8
GC86 (SEQ ID NO: 59)	CCTGAGCAGGATGGTGTT	8
RC99 (SEQ ID NO: 58)	CCTGAGCACGCTGGTATT	8

Results of PCRs with the primer pair NIT3 and 27f showed that the NIT3 primer specifically amplified only RC44 clone inserts (*Nitrobacter*) and not those from *Nitrospira* clones.

The different primer pairs were then used with DNAs extracted from sludges and the results are tabulated below in Table 8. The scorings presented in the table were generated by quantitating by eye the intensity of the amplificate in a stained gel. A definition of the scoring follows: - = no band; +/- = very faint band; + through ++++ = increasing intensity of the amplificate.

Table 8

Wastewater Treatment Plant	**Performance**	**MOS635r-27f 620 bp**	**NIT3-27f 1020 bp**
Oxley	Full nitrification	++++	++
Merrimac	Full nitrification	++++	++
Loganholme	Full nitrification	+++	+/-
Gibson Island	Full nitrification	+++	-
Fairfield	No nitrification	+/-	+++
Cannon Hill	Full nitrification	+	+
NOSBR	NO_2^- oxidation	+++++	++++
Saline waste water BNR SBR	Partial nitrification	+/-	++
Nitrifying biofilm reactor	Full nitrification	++++	++++
Phenol/cyanide removing SBR	No nitrification	+/-	++
BNR SBR	Full nitrification	+	+

These results show that in plants having good nitrification, *Nitraspira* species were present as evidenced by amplification of target DNA with the selected primer pairs.

REFERENCES

Blackall, L. L. (1994). Molecular identification of activated sludge foaming bacteria. *Water Science and Technology* **29-7**, 35-42.

Blackall, L. L., Seviour, E.M., Cunningham, M.A., Seviour, R.J., and Hugenholtz, P. (1994). "*Microthrix parvicella*" is a novel, deep branching member of the actinomycetes subphylum. *Systematic and Applied Microbiology* **17**, 513-518.

Blackburn, T. H. (1983). The Microbial Nitrogen Cycle. In *Microbial Geochemistry*, (pp. 63-89). Edited by W. E. Krumbein. Oxford: Blackwell Scientific Publications.

Bock, E., Koops, H., Ahlers, B., and Harms, H. (1992). Oxidation of inorganic nitrogen compounds as energy source. In *The Prokaryotes - A Handbook on the Biology of Bacteria: Ecophysiology, Isolation, Identification, Applications*, pp. 414-430. Edited by A. Balows, H. G. Trüper, M. Dworkin, W. Harder & K.-H. Schleifer. New York: Springer-Verlag.

Bond, P.L., Hugenholtz, P., Keller, J., and Blackall, L.L. (1995). Bacterial community structures of phosphate-removing and non-phosphate-removing activated sludges from sequencing batch reactors. *Applied and Environmental Microbiology*, **61**, 1910-1916.

Burrell, P., and Blackall, L.L. (in press). The microbiology of nitrogen removal in activated sludge systems. In *The Microbiology of Activated Sludge*, Edited by R. J. Seviour & L. L. Blackall.

Ehrich, S., Behrens, D., Lebedeva, E., Ludwig, W. and Bock, E. (1995). A new obligately chemolithoautotrophic, nitrite-oxidizing bacterium, *Nitrospira moscoviensis* sp. nov. and its phylogenetic relationship. *Archives of Microbiology*, **164**, 16-23.

Halling-Sørensen, B., and Jørgensen, S.E. (1993). *The Removal of Nitrogen Compounds from Wastewater*, Amsterdam: Elsevier.

Hovanec, T. A. & DeLong, E. F. (1996). Comparative Analysis of Nitrifying Bacteria Associated with Freshwater and Marine Aquaria. *Applied and Environmental Microbiology*, **62**, 2888-2896.

Meganck, M.T.J., and Faup, G.M. (1988). Enhanced biological phosphorus removal from waste waters. *Biotreatment Systems*, **3**, 111-204.

Randall, C. W. (1992). Introduction and Principles of Biological Nutrient Removal. In *Design and Retrofit of Wastewater Treatment Plants for Biological Nutrient Removal*, (pp. 7-84). Edited by C. W. Randall. Lancaster: Technomic Publishing Company Inc.

Robertson, L. A. & Kuenen, J. G. (1991). Physiology of Nitrifying and Denitrifying Bacteria. In *Microbial Production and Consumption of Greenhouse Gases: Methane, Nitrogen Oxides and Halomethanes*, (pp. 189-199). Edited by J. E. Rogers & W. B. Whitman. Washington D.C.: American Society for Microbiology.

Sambrook, J., Fritsch, E. F., and Maniatis, T., *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Plainview, New York, 1989.

Stackebrandt, E., and Goodfellow, M., eds, *Nucleic Acid Techniques in Bacterial Systematics*. John Wiley & Sons, New York, 1991

Wagner, M., Rath, G., Koops, H.-P., Flood, J., and Amann, R. (1996). In situ analysis of nitrifying bacteria in sewage treatment plants. *Water Science and Technology*, **34**, 237-244.

Weidner, S., Arnold, W. and Pühler, A. (1996). Diversity of Uncultured Microorganisms Associated with the seagrass *Halophila stipulacea* Estimated by Restriction Fragment Length Polymorphism Analysis of PCR-Amplified 16S rRNA Genes. *Applied and Environmental Microbiology*, **62**, 766-771.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: CRC for Waste Managment and Pollution Control Limited
(B) STREET: High Street
(C) CITY: Kensington
(D) STATE: New South Wales
(E) COUNTRY: Australia
(F) POSTAL CODE (ZIP): 2033

(ii) TITLE OF INVENTION: Aquatic Nitrite Oxidising Microorganisms

(iii) NUMBER OF SEQUENCES: 59

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1428 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
CAAGTCGAGC GAGAAGACGT AGCAATACGT TTGTAAAGCG GCGAACGGGT GAGGAATACA    60

TGGGTAACCT ACCTTCGAGT GGGGAATAAC TAGCCGAAAG GTTAGCTAAT ACCGCATACG   120

ACTCCTGGTC TGCGGATCGG GAGAGAAAGC GATACCGTGG GTATCGCGCT CTTGGATGGG   180

CTCATGTCCT ATCAGCTTGT TGGTGAGGTA ACGGCTCACC AAGGCTTCGA CGGGTAGCTG   240

GTCTGAGAGG ACGATCAGCC ACACTGGCAC TGCGACACGG GCCAGACTCC TACGGGAGGC   300

AGCAGTAAGG AATATTGCGC AATGGGCGAC AGCCTGACGC AGCNACGCCG CGTGGGGGAT   360
```

GAAGGTCTTC GGATTGTAAA CCCCTTTCGG CAGGGAAGAT GGAACGGGTA ACCGTTCGGA 420
CGGTACCTGC AGAAGCAGCC ACGGCTAACT TCGTGCCAGC AGCCGCGGTA ATACGAAGGT 480
GGCAAGCGTT GTTCGGATTT ACTGGGCGTA CAGGGAGCGT AGGCGGTTGG GTAAGCCCTC 540
CGTGAAATCT CCGGGCCTAA CCCGGAAAGT GCGGAGGGGA CTGCTCGGCT AGAGGATGGG 600
AGAGGAGCGC GGAATTCCCG GTGTAGCGGT GAAATGCGTA GAGATCGGGA GGAAGGCCGG 660
TGGCGAAGGC GGCGCTCTGG AACATTTCTG ACGCTGAGGC TCGAAAGCGT GGGGAGCAAA 720
CAGGATTAGA TACCCTGGTA GTCCACGCCT TAAACGATGG ATACTAAGTG TCGGCGGGTT 780
ACCGCCGGTG CCGCAGCTAA CGCATTAAGT ATCCCGCCTG GGAAGTACGG CCGCAAGGTT 840
GAAACTCAAA GGAATTGACG GGGGCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC 900
GCAACGCGAA GAACCTTACC CAGGCTGGAC ATGCAGGTAG TAGAAGGGTG AAAGCCTAAC 960
GAGGTAGCAA TACCATCCTG CTCAGGTGCT GCATGGCTGT CGTCAGCTCG TGCCGTGAGG 1020
TGTTGGGTTA AGTCCCGCAA CGAGCGCAAC CCCTGTCTTC AGTTACCAAC GGGTCATGCC 1080
GGGAACTCTG GAGAGACTGC CCAGGAGAAC GGGGAGGAAG GTGGGGATGA CGTCAAGTCA 1140
GCATGGCCTT TATGCCTGGG GCCACACACG TGCTACAATG GCCGGTACAA AGCGCTGCAA 1200
ACCCGTAAGG GGGAGCCAAT CCCAAAAAAC CGGCCTCAGT TCAGATTGAG GTCTGCAACT 1260
CGACCTCATG AAGGCGGAAT CGCTAGTAAT CCCGGATCAG CACGCCGGGG TGAATACGTN 1320
CCCGGGCCTT GTACACACCG CCCGTCACAC CACGAAAGTT TGTTGTACCT GAAGTCGTTG 1380
GCGCCAACCG CAAGGAGGCA GACGCCCACG GTATGACCGA TGATTGGG 1428

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1407 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
TAATACATGC AAGTCGAGCG AGAAGACGTA GCAATACGTT TGTAAAGCGG CGAACGGGTG   60
AGGAATACAT GGGTAGCCTA CCCTCGAGTG GGGAATAACT AACCGAAAGG TTAGCTAATA  120
CCGCATACGG CTCCTGGTCT GCGGATCGGG AGAGAAAGCG ATACCGTGGG TATCGCGCTC  180
TTGGATGGGC TCATGTCCTA TCAGCTTGTT GGTGAGGTAA CGGCTCACCA AGGCTTCGAC  240
GGGTAGCTGG TCTGAGAGGA CGATCAGCCA CACTGGCACT GCGACACGGG CCAGACTCCT  300
ACGGGAGGCA GCAGTAAGGA ATATTGCGCA ATGGGCGACA GCCTGACGCA GCNACGCCGC  360
GTGGGGGATG AAGGTCTTCG GATTGTAAAC CCCTTTCGGC AGGGAAGATG GAACGGGTAA  420
CCGTTCGGAC GGTACCTGCA GAAGCAGCCA CGGCTAACTT CGTGCCAGCA GCCGCGGTAA  480
TACGAAGGTG GCAAGCGTTG TTCGGATTTA CTGGGCGTAC AGGGAGCGTA GGCGGTTGGG  540
TAAGCCCTCC GTGAAATCTC CGGGCCTAAC CCGGAAAGTG CGGAGGGGAC TGCTCGGCTA  600
GAGGATGGGA GAGGAGCGCG GAATTCCCGG TGTAGCGGTG AAATGCGTAG AGATCGGGAG  660
GAAGGCCGGT GGCGAAGGCG GCGCTCTGGA ACATTTCTGA CGCTGAGGCT CGAAAGCGTG  720
GGGAGCAAAC AGGATTAGAT ACCCTGGTAG TCCACGCCTT AAACGATGGA TACTAAGTGT  780
CGGCGGGTTA CCGCCGGTGC CGCAGCTAAC GCATTAAGTA TCCCGCCTGG GAAGTACGGC  840
CGCAAGGTTG AAACTCAAAG GAATTGACGG GGGCCCGCAC AAGCGGTGGA GCATGTGGTT  900
TAATTCGACG CAACGCGAAG AACCTTACCC AGGCTGGACA TGCAGGTAGT AGAAGGGTGA  960
AAGCCTAACG AGGTAGCAAT ACCATCCTGC TCAGGTGCTG CATGGCTGTC GTCAGCTCGT 1020
GCCGTGAGGT GTTGGGTTAA GTCCCGCAAC GAGCGCAACC CCTGTCTTCA GTTACCAACG 1080
GGTCATGCCG GGAACTCTGG AGAGACTGCC CAGGAGAACG GGGGAGGAAG GTGGGGATGA 1140
CGTCAAGTCA GCATGGCCTT TATGCCTGGG GCCACACACG TGCTACAATG GCCGGTACAA 1200
AGCGCTGCAA ACCCGTAAGG GGGAGCCAAT CGCAAAAAAC CGGCCTCAGT TCAGATTGAG 1260
GTCTGCAACT CGACCTCATG AAGGCGGAAT CGCTAGTAAT CCCGGATCAG CACGCCGGGG 1320
TGAATACGTN CCCGGACCTT GTACACACCG CCCGTCACAC CACGAAAGTT TGTTGTACCT 1380
GAAGTCGTTG GCGCCAACCG CAAGGAG                                     1407
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1500 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
TTGATCCTGG CTCAGAACGA ACGCTGGCGG CGCGCCTAAT ACATGCAAGT CGAGCGAGAA     60
GACGTAGCAA TACGTTTGTA AAGCGGCGAA CGGGTGAGGA ATACATGGGT AACCTACCCT    120
CGAGTGGGGA ATAACTAGCC GAAAGGTTAG CTAATACCGC ATACGACTCC TGGTCTGCGG    180
ATCGGGAGAG AAAGCGATAC CGTGGGTATC GCGCTCTTGG ATGGGCTCAT GTCCTATCAG    240
CTTGTTGGTG AGGTAACGGC TCACCAAGGC TTCGACGGGT AGCTGGTCTG AGAGGACGAT    300
CAGCCACACT GGCACTGCGA CACGGGCCAG ACTCCTACGG GAGGCAGCAG TAAGGAATAT    360
TGCGCAATGG GCGACAGCCT GACGCAGCNA CGCCGCGTGG GGGATGAAGG TCTTCGGATT    420
GTAAACCCCT TTCGGCAGGG AAGATGGAAC GGGTAACCGT TCGGACGGTA CCTGCAGAAG    480
CAGCCACGGC TAACTTCGTG CCAGCAGCCG CGGTAATACG AAGGTGGCAA GCGTTGTTCG    540
GATTTACTGG GCGTACAGGG AGCGTAGGCG GTTGGGTAAG CCCTCCGTGA AATCTCCGGG    600
CCTAACCCGG AAAGTGCGGA GGGGACTGCT CGGCTAGAGG ATGGGAGAGG AGCGCGGAAT    660
TCCCGGTGTA GCGGTGAAAT GCGTAGAGAT CGGGAGGAAG GCCGGTGGCG AAGGCGGCGC    720
TCTGGAACAT TTCTGACGCT GAGGCTCGAA AGCGTGGGGA GCAAACAGGA TTAGATACCC    780
TGGTAGTCCA CGCCTTAAAC GATGGATACT AAGTGTCGGC GGGTTACCGC CGGTGCCGCA    840
GCTAACGCAT TAAGTATCCC GCCTGGGAAG TACGGCCGCA AGGTTGAAAC TCAAAGGAAT    900
TGACGGGGGC CCGCACAAGC GGTGGAGCAT GTGGTTTAAT TCGACGCAAC GCGAAGAACC    960
TTACCCAGGC TGGACATGCA GGTAGTAGAA GGGTGAAAGC CTAACGAGGT AGCAACACCA   1020
TCCTGCTCAG GTGCTGCATG GCTGTCGTCA GCTCGTGCCG TGAGGTGTTG GGTTAAGTCC   1080
CGCAACGAGC GCAACCCCTG TCTTCAGTTA CCAACGGGTC ATGCCGGGAA CTCTGGAGAG   1140
ACTGCCCAGG AGAACGGGGA GGAAGGTGGG GATGACGTCA AGTCAGCATG GCCTTTATGC   1200
CTGGGGCCAC ACACGTGCTA CAATGGCCGG TACAAAGCGC TGCAAACCCG TAAGGGGGAG   1260
CCAATCGCAA AAAACCGGCC TCAGTTCAGA TTGAGGTCTG CAACTCGACC TCATGAAGGC   1320
```

GGAATCGCTA GTAATCCCGG ATCAGCACGC CGGGGTGAAT ACGTNCCCGG GCCTTGTACA 1380

CACCGCCCGT CACACCACGA AAGTTTGTTG TACCTGAAGT CGTTGGCGCC AACCGCAAGG 1440

GGGCAGACGC CCACGGTATG ACCGATGATT GGGGTGAAGT CGTAACAAGG TAACCGTAAC 1500

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1420 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

CGAGAAGACG TAGCAATACG TTTGTAAAGC GGCGAACGGG TGAGGAATAC ATGGGTAACC 60

TACCCTCGAG TGGGGAATAA CTAACCGAAA GGTTAGCTAA TACCGCATAC GGCTCCTGGT 120

CTGCGGATCG GGAGAGAAAG CGATACCGTG GGTATCGCGC TCTTGGATGG GCTCATGTCC 180

TATCAGCTTG TTGGTGAGGT AACGGCTCAC CAAGGCTTCG ACGGGTAGCT GGTCTGAGAG 240

GACGATCAGC CACACTGGCA CTGCGACACG GGCCAGACTC CTACGGGAGG CAGCAGTAAG 300

GAATATTGCG CAATGGGCGA CAGCCTGACG CAGCGACGCC GCGTTGGGGA TGAAAGTCTT 360

CCGATTGTAA ACCCCTTTCC GCAGGGAAGA TGGAACGGGT AACCGTTCGG ACGGTACCTG 420

CAGAAGCAGC CACGGCTAAC TTCGTGCCAG CAGCCGCGGT AATACGAAGG TGGCAAGCGT 480

TGTTCGGATT TACTGGGCGT ACAGGGAGCG TAGGCGGTTG GGTAAGCCCT CCGTGAAATC 540

TCCGGGCCTA ACCCGGAAAG TGCGGAGGGG ACTGCTCGGC TAGAGGATGG GAGAGGAGCG 600

CGGAATTCCC GGTGTAGCGG TGAAATGCGT AGAGATCGGG AGGAAGGCCG GTGGCGAAGG 660

CGGCGCTCTG GAACATTTCT GACGCTGAGG CTCGAAAGCG TGGGGAGCAA ACAGGATTAG 720

ATACCCTGGT AGTCCACGCC TTAAACGATG GATACTAAGT GTCGGCGGGT TACCGCCGGT 780

GCCGCAGCTA ACGCATTAAG TATCCCGCCT GGGAAGTACG GCCGCAAGGT TGAAACTCAA 840

AGGAATTGAC GGGGCCCCGC ACAAGCGGTG GAGCATGTGG TTTAATTCGA CGCAACGCGA 900

AGAACCTTAC CCAGGCAGGA CATGCAGGTA GTAGAAGGGT GAAAGCCTAA CGAGGTAGCA 960

ATACCATCCT GCTCAGGTGC TGCATGGCTG TCGTCAGCTC GTGCCGTGAG GTGTTGGGTT 1020

AAGTCCCGCA ACGAGCGCAA CCCCTGTCTT CAGTTACCAA CGGGTCATGC CGGGAACTCT 1080

GGAGAGACTG CCCAGGAGAA CGGGGAGGAA GGTGGGGATG ACGTCAAGTC AGCATGGCCT 1140

TTATGCCTGG GGCCACACAC GTGCTACAAT GGCCGGTACA AAGCGCTGCA AACCCGTAAG 1200

GGGGAGCCAA TCGCAAAAAA CCGGCCTCAG TTCAGATTGA GGTCTGCAAC TCGACCTCAT 1260

GAAGGCGGAA TCGCTAGTAA TCCCGGATCA GCACGCCGGG GTGAATACGT NCCCGGGCCT 1320

TGTACACACC GCCCGTCACA CCACGAAAGT TTGTTGTACC TGAAGTCGTT GGCGCCAACC 1380

GCAAGGAGGC AGACGCCCAC GGTATGACCG ATGATTGGGG 1420

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1505 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

AGAGTTTGAT CCTGGCTCAG AACGAACGCT GGCGGCGCGC CTAATACATG CAAGTCGAGC 60

GAGAAGACGT AGCAATACGT TTGTAAAGCG GCGAACGGGT GAGGAATACA TGGGTAATCT 120

ACCATCGAGT GGGGAATAAC CAACCGAAAG GTTGGCTAAT ACCGCGTACG CTTCTGAGTC 180

TTCGGGTTCG GAAGGAAAGC CGTACTGTGA GTGCGGCGCT CTTTGATGAG CTCATGTCCT 240

ATCAGCTTGT TGGTAGGGTA ACGGCCTACC AAGGCTTTGA CGGGTAGCTG GTCTGAGAGG 300

ACGATCAGCC ACACTGGCAC TGCGACACGG GCCAGACTCC TACGGGAGGC AGCAGTAAGG 360

AATATTGCGC AATGGGCGAA AGCCTGACGC AGCNACGCCG CGTGGGGGAT GAAGGTCTTC 420

GGATTGTAAA CCCCTTTCGG GAGGGAAGAT GGAGCGAGCA ATCGTTCGGA CGGTACCTCC 480

AGAAGCAGCC ACGGCCAACT TCGTGCCAGC AGCCGCGGTA ATACGAAGGT GGCAAGCGTT 540

GTTCGGATTC ACTGGGCGTA CAGGGTGTGT AGGCGGTTTG GTAAGCCTTC TGTTAAAGCT 600

TCGGGCCCAA CCCGGAAAGC GCAGACGGTA CTGCCAGGCT AGAGGGTGGG AGAGGAGCGC 660

GGAATTCCCG GTGTAGCGGT GAAATGCGTA GAGATCGGGA GGAAGGCCGG TGGCGAAGGC 720

GGCGCTCTGG AACATACCTG ACGCTGAGAC ACGAAAGCGT GGGGAGCAAA CAGGATTAGA 780

TACCCTGGTA GTCCACGCCC TAAACTATGG ATACTAAGTG TCGGCGGGTT ACCGCCGGTG 840

CCGCAGCTAA CGCATTAAGT ATCCCGCCTG GGAAGTACGG CCGCAAGGTT GAAACTCAAA 900

GGAATTGACG GGGGCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC GCAACGCGAA 960

GAACCTTACC CAGGTTGGAC ATGCACGTAG TAGAAAGGTG AAAGCCTGAC GAGGTAGCAA 1020

TACCAGCGTG CTCAGGTGCT GCATGGCTGT CGTCAGCTCG TGCCGTGAGG TGTTGGGTTA 1080

AGTCCCGCAA CGAGCGCAAC CCCTGCTTTC AGTTGCTACC GGGTCATGCC GAGCACTCTG 1140

AAAGGACTGC CCAGGATAAC GGGGAGGAAG GTGGGGATGA CGTCAAGTCA GCATGGCCTT 1200

TATGCCTGGG GCCACACACG TGCTACAATG GCCGGTACAA AGCGCTGCAA ACCCGTGAGG 1260

GGGAGCCAAT CGCAAAAAAC CGGCCTCAGT TCAGATTGAG GTCTGCAACT CGACCTCATG 1320

AAGGCGGAAT CGCTAGTAAT CGCGGATCAG CACGCCGCGG TGAATACGTN CCCGGGCCTT 1380

GTACACACCG CCCGTCACAC CACGAAAGCC TGTTGTACCT GAAGTCGCCC AAGCCAACCG 1440

CAAGGAGGCA GGCGCCCACG GTATGGCCCG TGATTGGGGT GAAGTCGTAA CAAGGTAACC 1500

GTAAA 1505

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1441 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

AAGTCGAGCG AGAAGGTGTA GCAATACACT TGTAAAGCGG CGAACGGGTG AGGAATACAT 60

```
GGGTAATCTA CCATCGAGTG GGGAATAACC AGCCGAAAGG TTGGCTAATA CCGCGTACGC     120
TTCCGAGTCT TCGGGCTTGG AAGGAAAGCC GCACTGTGAG TGCGGCGCTC TTTGATGAGC     180
TCATGTCCTA TCAGCTTGTT GGTAGGGTAA CGGCCTACCA AGGCTTTGAC GGGTAGCTGG     240
TCTGAGAGGA CGATCAGCCA CACTGGCACT GCGACACGGG CCAGACTCCT ACGGGAGGCA     300
GCAGTAAGGA ATATTGCGCA ATGGGCGAAA GCCTGACGCA GCGACGCCGC GTGGGGGATG     360
AAGGTCTTCG GATTGTAAAC CCCTTTCGGG AGGGAAGATG GAGCCAGCAA TCGTTCGGAC     420
GGTACCTCCA GAAGCAGCCA CGGCCAACTT CGTGCCAGCA GCCGCGGTAA TACGAAGGTG     480
GCAAGCGTTG TTCGGATTCA CTGGGCGTAC AGGGTGTGTA NGCGGTTTGG TAAGCCTTCT     540
GTTAAAGCTT CGGGCCCAAC CCGGAAAGCG CAGAGGGTAC TGCCAGGCTA GAGGGTGGGA     600
GAGGAGCGCG GAATTCCCGG TGTAGCGGTG AAATGCGTAG AGATCGGGAG GAAGGCCGGT     660
GGCGAAGGCG GCGCTCTGGA ACATGCCTGA CGCTGAGACA CGAAAGCGTG GGGAGCAAAC     720
AGGATTAGAT ACCCTGGTAG TCCACGCCCT AAACTATGGA TACTAAGTGT CGGCGGGTTA     780
CCGCCGGTGC CGCAGCTAAC GCATTAAGTA TCCCGCCTGG GAAGTACGGC CGCAAGGTTG     840
AAACTCAAAG GAATTGACGG GGGCCCGCAC AAGCGGTGGA GCATGTGGTT TAATTCGACG     900
CAACGCGAAG AACCTTACCC AGGTTGGACA TGCACGTAGT AGAAAGGTGA AAGNCTAACG     960
AGGTAGCAAT ACCAGCGTGC TCAGGTGCTG CATGGCTGTC GTCAGCTCGT GCCGTGAGGT    1020
GTTGGGTTAA GTCCCGCAAC GAGCGCAACC CCTGCTTTCA GTTGCTACCG GGTCATGCCG    1080
AGCACTCTGA AAGGACTGCC CAGGATAACG GGGAGGAAGG TGGGGATGAC GTCAAGTCAG    1140
CATGGCCTTT ATGCCTGGGG CCACACACGT GCTACAATGG CCGGTACAAA GCGCTGCAAA    1200
CCCGTGAGGG GGAGCCAATC GCAAAAAACC GGCCTCAGTT CAGATTGAGG TCTGCAACTC    1260
GACCTCATGA AGGCGGAATC GCTAGTAATC GCGGATCAGC ACGCCGCGGT GAATACGTNC    1320
CCGGGCCTTG TACACACCGC CCGTCACACC ACGAAAGCCT GTTGTACCTG AAGTCGCCCA    1380
AGCCAACCGC AAGGAGGCAG GCGCCCACGG TATGGCCGGT GATTGGGGTG AAGTCCTAAC    1440
A                                                                    1441
```

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1426 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```
TAATACATGC AAGTCGAGCG AGAAGGTGTA GCAATACACT TGTAAAGCGG CGAACGGGTG     60
AGGAATACAT GGGTAATCTA CCATCGAGTG GGGAATAACC AACCGAAAGG TTGGCTAATA    120
CCGCGTACGC TTCTGAGCCT TCGTGTTCGG AAGGAAAGCC GTACTGTGAG TGCGGCGCTC    180
TTTGATGAGC TCATGTCCTA TCAGCTTGTT GGTAGGGTAA CGGCCTACCA AGGCTTTGAC    240
GGGTAGCTGG TCTGAGAGGA CGATCAGCCA CACTGGCACT GCGACACGGG CCAGACTCCT    300
ACGGGAGGCA GCAGTAAGGA ATATTGCGCA ATGGGCGAAA GCCTGACGCA GCNACGCCGC    360
GTGGGGGATG AAGGTCTTCG GATTGTAAAC CCCTTTCGGG AGGGAAGATG GAGCGAGCAA    420
TCGTTCGGAC GGTACCTCCA GAAGCAGCCA CGGCCAACTT CGTGCCAGCA GCCGCGGTAA    480
TACGAAGGTG GCAAGCGTTG CTTGGATTCA CTGGGCGTAC AGGGTGTGTA GGCGGTTTGG    540
TAAGCCTTCT GTTAAAGCTT CGGGCCCAAC CCGAAAAGCG CAGAGGGTAC TGCCAGGCTA    600
GAGGGTGGGA GAGGAGCGCG GAATTCCCGG TGTAGCGGTG AAATGCGTAG AGATCGGGAG    660
GAAGGCCGGT GGCGAAGGCG GCGCTCTGGA ACATACCTGA CGCTGAGACA CGAAAACGTG    720
GGGAGCAAAC AGGATTAGAT ACCCTGGTAG TCCACGCCCT AAACTATGGA TACTAAGTGT    780
CGGCGGGTTA CCGCCGGTGC CGCAGCTAAC GCATTAAGTA TCCCGCCTGG GAGGTACGGC    840
CGCAAGGTTG AAACTCAAAG GAATTGACGG GGGCCCGCAC AAGCGGTGGA GCTTGTGGTT    900
TAATTCGACG CAACGCGAAG AACCTTACCC AGGTTGGACA TGCACGTAGT AGAAAGGTGA    960
AAGCCTGACG AGGTAGCAAT ACCAGCGTGC TCAGGTGCTG CATGGCTGTC GTCAGCTCGT   1020
GCCGTGAGGT GTTGGGTTAA GTCCCGCAAC GAGCGCAACC CCTGCTTTCA GTTGCTACCG   1080
GGTCATGCCG AGCACTCTGA AAGGACTGCC CAGGATAACG GGGAGGAAGG TGGGGATGAC   1140
GTCAAGTCAG CATGGCCTTT ATGCCTGGGG CCACACACGT GCTACAATGG CCGGTACAAA   1200
GCGCTGCAAA CCCGTGAGGG GGAGCCAATC GCAAAAAACC GGCCTCAGTT CAGATTGAGG   1260
TCTGCAACTC GACCTCATGA AGGCGGAATC GCTAGTAATC GCGGATCAGC ACGCCGCGGT   1320
```

GAATACGTNC CCGGGCCTTG TACACACCGC CCGTCACACC ACGAAAGCCT GTTGTACCTG 1380

AAGTCGCCCA AGCCAACCGC AAGGAGGCAG GCGCCCACGG TATGGC 1426

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1429 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

TAATACATGC AAGTCGAGCG AGAAGGTGTA GCAATACACT TGTAAAGCGG CGAACGGGTG 60

AGGAATACAT GGGTAATCTA CCATCGAGTG GGGAATAACC AACCGAAAGG TTGGCTAATA 120

CCGCGTACGC CTCCGAGTCT TCGGGTTCGG AGGGAAAGCT GCACTGTGAG TGTAGCGCTC 180

TTTGATGAGC TCATGTCCTA TCAGCTTGTT GGTAGGGTAA CGGCCTACCA AGGCTTTGAC 240

GGGTAGCTGG TCTGAGAGGA CGATCAGCCA CACTGGCACT GCGACACGGG CCAGACTCCT 300

ACGGGAGGCA GCAGTAAGGA ATATTGCGCA ATGGGCGAAA GCCTGACGCA GCNACGCCGC 360

GTGGGGGATG AAGGTCTTCG GATTGTAAAC CCCTTTCGGG AGGGAAGATG GAGCGAGCAA 420

TCGTTCGGAC GGTACCTCCA GAAGCAGCCA CGGCCAACTT CGTGCCAGCA GCCGCGGTAA 480

TACGAAGGTG GCAAGCGTTG TTCGGATTCA CTGGGCGTAC AGGGTGTGTA GGCGGTTTGG 540

TAAGCCTTCT GTTAAAGCTT CGGGCCCAAC CCGGAAAGCG CAGGGGGTAC TGCCAGGCTA 600

GAGGGTGGGA GAGGAGCGCG GAATTCCCGG TGTAGCGGTG AAATGCGTAG AGATCGGGAG 660

GAAGGCCGGT GGCGAAGGCG GCGCTCTGGA ACATACCTGA CGCTGAGACA CGAAAGCGTG 720

GGGAGCAAAC AGGATTAGAT ACCCTGGTAG TCCACGCCCT AAGCTATGGA TACTAAGTGT 780

CGGCGGGTTA CCGCCGGTGC CGCAGCCAAC GCGTTAAGTA TCCCGCCTGG GAAGTACGGC 840

CGCAAGGTTG AAACTCAAAG GAATTGACGG GGGCCCGCAC AAGCGGTGGA GCATGTGGTT 900

TAATTCGACG CAACGCGAAG AACCTTACCC AGGTTGGACA TGCACGTAGT AGAAAGGTGA 960

AAGCCTGACG AGGTAGCAAT ACCAGCGTGC TCAGGTGCTG CATGGCTGTC GTCAGCTCGT 1020

GCCGTGAGGT GTTGGGTTAA GTCCCGCAAC GAGCGCAACC CCTGCTTTCA GTTGCTACCG 1080

GGTCATGCCG AGCACTCTGA AAGGACTGCC CAGGATAACG GGGGAGGAAG GTGGGGATGA 1140

CGTCAAGTCA GCATGGCCTT TATGCCTGGG GCCACACACG TGCTACAATG GCCGGTACAA 1200

AACGCTGCAA ACCCGTGAGG GGGAGCCAAT CGCAAAAAAC CGGCCTCAGT TCAGATTGAG 1260

GTCTGCAACT CGACCTCATG AAGGCGGAAT CGCTAGTAAT CGCGGATCAG CACGCCGCGG 1320

TGAATACGTN CCCGGGCCTT GTGCACACCG CCCGTCACAC CACGAAAGCC TGTTGTACCT 1380

GAAGTCGCCC AAGCCAACCG CAAGGAGGCA GGCGCCCACG GTATGGCCG 1429

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1415 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CGAGAAGGTG TAGCAATACA CTTGTAAAGC GGCGAACGGG TGAGGAATAC ATGGGTAATC 60

TACCATCGAG TGGGGAATAA CCAACCGAAA GGTTGGCTAA TACCGCGTAC GCCTCCGAGT 120

CTTCGGGTTC GGAGGGAAAG CTGCACTGTG AGTGTAGCGC TCTTTGATGA GCTCATGTCC 180

TATCAGCTTG TTGGTAGGGT AACGGCCTAC CAAGGCTTTG ACGGGTAGCT GGTCTGAGAG 240

GACGATCAGC CACACTGGCA CTGCGACACG GGCCAGACTC CTACGGGAGG CAGCAGTAAG 300

GAATATTGCG CAATGGGCGA AAGCCTGACG CAGCNACGCC GCGTGGGGGA TGAAGGTCTT 360

CGGATTGTAA ACCCCTTTCG GGAGGGAAGA TGGAGCGAGC AATCGTTCGG ACGGTACCTC 420

CAGAAGCAGC CACGGCCAAC TTCGTGCCAG CAGCCGCGGT AATACGAAGG TGGCAAGCGT 480

TGTTCGGATT CACTGGGCGT ACAGGGTGTG TAGGCGGTTT GGTAAGCCTT CTGTTAAAGC 540

TTCGGGCCCA ACCCGGAAAG CGCAGAGGGT ACTGCCAGGC TAGAGGGTGG GAGAGGAGCG 600

CGGAATTCCC GGTGTAGCGG TGAAATGCGT AGAGATCGGG AGGAAGGCCG GTGGCGAAGG 660

CGGCGCTCTG GAACATACCT GACGCTGAGA CACGAAAGCG TGGGGAGCAA ACAGGATTAG 720

ATACCCTGGT AGTCCACGCC CTAAACTATG GATACTAAGT GTCGGCGGGT TACCGCCGGT 780

GCCGCAGCTA ACGCATTAAG TATCCCGCCT GGGAAGTACG GCCGCAAGGT TGAAACTCAA 840

AGGAATTGAC GGGGGCCCGC ACAAGCGGTG GAGCATGTGG TTTAATTCGA CGCAACGCGA 900

AGAACCTTAC CCAGGTTGGA CATGCACGTA GTAGAAAGGT GAAAGCCTGA CGAGGTAGCA 960

ATACCAGCGT GCTCAGGTGC TGCATGGCTG TCGTCAGCTC GTGCCGTGAG GTGTTGGGTT 1020

AAGTCCCGCA ACGAGCGCAA CCCCTGCTTT CAGTTGCTAC CGGGTCATGC CGAGCACTCT 1080

GAAAGGACTG CCCAGGATAA CGGGGAGGAA GGTGGGGATG ACGTCAAGTC AGCATGGCCT 1140

TTATGCCTGG GGCCACACAC GTGCTACAAT GGCCGGTATA AAACGCTGCA AACCCGTGAG 1200

GGGGAGCCAA TCGCAAAAAA CCGGCCTCAG TTCAGATTGA GGTCTGCAAC TCGACCTCAT 1260

GAAGGCGGAA TCGCTAGTAA TCGCGGATCA GCACGCCGCG GTGAATACGT NCCCGGGCCT 1320

TGTACACACC GCCCGTCACA CCACGAAAGC CTGTTGTACC TGAAGTCGCC CAAGCCAACC 1380

GCAAGGAGGC AGGCGCCCAC GGTATGGCCG GTGAT 1415

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1435 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CCTAATACAT GCAAGTCGAT CGAGAAGGTG TAGCAATACA CTTGTAAAGC GGCGAACGGG 60

TGAGGAATAC ATGGGTAATC TACCATCGAG TGGGGAATAA CCAACCGAAA GGTTGGCTAA 120

TACCGCGTAC GCCTCCGAGT CTTCGGGTTC GGAGGGAAAG CTGCACTGTG AGTGTAGCGC 180

TCTTTGATGA GCTCATGTCC TATCAGCTTG TTGGTAGGGT AACGGCCTAC CAAGGCTTTG 240

ACGGGTAGCT GGTCTGAGAG GACGATCAGC CACACTGGCA CTGCGACACG GGCCAGACTC 300

CTACGGGAGG CAGCAGTAAG GAATATTGCG CAATGGGCGA AAGCCTGACG CAGCCACGCC 360

GCGTGGGGGA TGAAGGTCTT CGGATTGTAA ACCCCTTTCG GGAGGGAAGA TGGAGCGAGC 420

AATCGTTCGG ACGGTACCTC CAGAAGCAGC CACGGCCAAC TTCGTGCCAG CAGCCGCGGT 480

AATACGAAGG TGGCAAGCGT TGTTCGGATT CACTGGGCGT ACAGGGTGTG TAGGCGGTTT 540

GGTAAGCCTT CTGTTAAAGC TTCGGGCCCA ACCCGGAAAG CGCAGAGGGT ACTGCCAGGC 600

TAGAGGGTGG GAGAGGAGCG CGGAATTCCC GGTGTAGCGG TGAAATGCGT AGAGATCGGG 660

AGGAAGGCCG GTGGCGAAGG CGGCGCTCTG GAACATACCT GACGCTGAGA CACGAAAGCG 720

TGGGGAGCAA ACAGGATTAG ATACCCTGGT AGTCCACGCC CTAAACTATG GATACTAAGT 780

GTCGGCGGGT TACCGCCGGT GCCGCAGCTA ACGCATTAAG TATCCCGCCT GGGAAGTACG 840

GCCGCAAGGT TGAAACTCAA AGGAATTGAC GGGGGCCCGC ACAAGCGGTG GAGCATGTGG 900

TTTAATTCGA CGCAACGCGA AGAACCTTAC CCAGGTTGGA CATGCACGTA GTAGAAAGGT 960

GAAAGCCTGA CGAGGTAGCA ATACCAGCGT GCTCAGGTGC TGCATGGCTG TCGTCAGCTC 1020

GTGCCGTGAG GTGTTGGGTT AAGTCCCGCA ACGAGCGCAA CCCCTGCTTT CAGTTGCTAC 1080

CGGGTCATGC CGAGCACTCT GAAAGGACTG CCCAGGATAA CGGGGAAGGA AGGTGGGGAT 1140

GACGTCAAGT CAGCATGGCC TTTATGCCTG GGGCCACACA CGTGCTACAA TGGCCGGTAC 1200

AAAACGCTGC AAACCCGTGA GGGGGAGCCA ATCGCAAAAA ACCGGCCTCA GTTCAGATTG 1260

AGGTCTGCAA CTCGACCTCA TGAAGGCGGA ATCGCTAGTA ATCGCGGATC AGCACGCCGC 1320

GGTGAATACG TNCCCGGGCC TTGTACACAC CGCCCGTCAC ACCACGAAAG CCTGTTGTAC 1380

CTGAAGTCGC CCAAGCCAAC CGCAAGAAGG CAGGCGCCCA CGGTATGGCC GGTGA 1435

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1437 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

```
AATACATGCA AGTCGATCGA GAAGGTGTAG CAATACACTT GTAAAGCGGC GAACGGGTGA     60
GGAATACATG GGTAATCTAC CATCGAGTGG GGAATAACCA ACCGAAAGGT TGGCTAATAC    120
CGCGTACGCC TCCGAGTCTT CGGGTTCGGA GGGAAAGCTG CACTGTGAGT GTAGCGCTCT    180
TTGATGAGCT CATGTCCTAT CAGCTTGTTG GTAGGGTAAC GGCCTACCAA GGCTTTGACG    240
GGTAGCTGGT CTGAGAGGAC GATCAGCCAC ACTGGCACTG CGACACGGGC CAGACTCCTA    300
CGGGAGGCAG CAGTAAGGAA TATTGCGCAA TGGGCGAAAG CCTGACGCAG CCACGCCGCG    360
TGGGGGATGA AGGTCTTCGG ATTGTAAACC CCTTTCGGGA GGGAAGATGG AGCGAGCAAT    420
CGTTCGGACG GTACCTCCAG AAGCAGCCAC GGCCAACTTC GTGCCAGCAG CCGCGGTAAT    480
ACGAAGGTGG CAAGCGTTGT TCGGATTCAC TGGGCGTACA GGGTGTGTAG GCGGTTTGGT    540
AAGCCTTCTG TTAAAGCTTC GGGCCCAACC CGGAAAGCGC AGAGGGTACT GCCAGGCTAG    600
AGGGTGGGAG AGGAGCGCGG AATTCCCGGT GTAGCGGTGA AATGCGTAGA GATCGGGAGG    660
AAGGCCGGTG GCGAAGGCGG CGCTCTGGAA CATACCTGAC GCTGAGACAC GAAAGCGTGG    720
GGAGCAAACA GGATTAGATA CCCTGGTAGT CCACGCCCTA AACTATGGAT ACTAAGTGTC    780
GGCGGGTTAC CGCCGGTGCC GCAGCTAACG CATTAAGTAT CCCGCCTGGG AAGTACGGCC    840
GCAAGGTTGA AACTCAAAGG AATTGACGGG GGCCCGCACA AGCGGTGGAG CATGTGGTTT    900
AATTCGACGC AACGCGAAGA ACCTTACCCA GGTTGGACAT GCACGTAGTA NAAAGGTGAA    960
AGCCTGACGA GGTAGCAATA CCAGCGTGCT CAGGTGCTGC ATGGCTGTCT TCAGCTCGTG   1020
CCGTGAGGTG TTGGGTTAAG TCCCGCAACG AGCGCAACCC CTGCTTTCAG TTGCTACCGG   1080
GTCATGCCGA ACACTCTGAA AGGACTGCCC AGGATAACGG GGAAGGAAGG TGGGGATGAC   1140
GTCAAGTCAG CATGGCCTTT ATGCCTGGGG CCACACACGT GCTACAATGG CCGGTACAAA   1200
GCGCTGCAAA CCCGTGAGGG GGAGCCAATC GCAAAAAACC GGCCTCAGTT CAGATTGAGG   1260
TCTGCAACTC GACCTCATGA AGGCGGAATC GCTAGTAATC GCGGATCAGC ACGCCGCGGT   1320
GAATACGTNC CCGGGCCTTG TACACACCGC CCGTCACACC ACGAAAGCCT GTTGTACCTG   1380
AAGTCGCCCA AGCCAACCGC AAGGAGGCAG GCGCCCACGG TATGGCCGGT GATGGGG      1437
```

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1437 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

```
AATACATGCA AGTCGATCGA NAAGGTGTAG CAATACACTT GTAAAGCGGC GAACGGGTGA     60
GGAATACATG GGTAATCTAC CATCGAGTGG GGAATAACCA ACCGAAAGGT TGGCTAATAC    120
CGCGTACGCC TCCGAGTCTT CGGGTTCGGA GGGAAAGCTG CACTGTGAGT GTAGCGCTCT    180
TTGATGAGCT CATGTCCTAT CAGCTTGTTG GTAGGGTAAC GGCCTACCAA GGCTTTGACG    240
GGTATCTGGT CTGAGAGGAC GATCAGCCAC ACTGGCACTG CGACACGGGC CAGACTCCTA    300
CGGGAGGCAG CAGTAAGGAA TATTGCGCAA TGGGCGAAAC CCNGACGCAG CCACGCCGCG    360
TGGGGGATGA AGGTCTTCGG ATTGTAAACC CCTTTCGGGA GGGAAGATGG AACGAGCAAT    420
CGTTCGGACG GTACCTCCAG AAGCAGCCAC GGCCAACTTC GTGCCAGCAG CCGCGGTAAT    480
ACGAAGGTGG CAAGCGTTGT TCGGATTCAC TGGGCGTACA GGGTGTGTAG GCGGTTTGGT    540
AAGCCTTCTG TTAAAGCTTC GGGCCCAACC CGGAAAGCGC AGAGGGTACT GCCAGGCTAG    600
AGGGTGGGAG AGGAGCGCGG AATTCCCGGT GTAGCGGTGA AATGCGTAGA GATCGGGAGG    660
AAGGCCGGTG GCGAAGGCGG CGCTCTGGAA CATACCTGAC GCTGAGACAC GAAAGCGTGG    720
GGNGCAAACA GGATTAGATA CCCTGGTAGT CCACGCCCTA AACTATGGAT ACTAAGTGTC    780
GGCGGGTTAC CGCCGGTGCC GCAGCTAACG CATTAAGTAT CCCGCCTGGG AAGTACGGCC    840
GCAAGGTTGA AACTCAAAGG GATTGACGGG GGCCCGCACA AGCGGTGGGG CATGTGGTTT    900
AATTCGACGC AACGCGAAGA ACCTTACCCA GGTTGGACAT GCACGTAGTN GAAAGGTGAA    960
AGCCTGACGA GGTAGCAATA CCAGCGTGCT CAGGTGCTGC ATGGCTGTCG TCAGCTCGTG   1020
CCGTGAGGTG TTGGGTTAAG TCCCGCAACG AGCGCAACCC CTGCTTTCAG TTGCTACCGG   1080
GTCATGCCGA ACACTCTGAA AGGACTGCCC AGGATAACGG GGAAGGAAGG TGGGGATGAC   1140
GTCAAGTCAG CATGGCCTTT ATACCTGGGG CCACACACGT GCTACAATGG CCGGTACAAA   1200
ACGCTGCAAA CCCGTGAGGG GGAGCCAATC GCAAAAAACC GGCCTCAGTT CAGATTGAGG   1260
```

TCTGCAACTC GACCTCATGA ATGCGGAATC GCTAGTAATC GCGGATCAGC ACGCCGCGGT 1320

GAATACGTNC CCGGGCCTTG TACACACCGC CCGTCACACC ACGAAAGCCT GTTGTACCTG 1380

AAGTCGCCCA AGCCAACCGC AAGGAGGCAG GCGCCCACGG TATGGCCGGT GATGGGG 1437

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1435 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TAATACATGC AAGTCGATCG ANAAGGTGTA GCAATACACT TGTAAAGCGG CGAACGGGTG 60

AGGAATACAT GGGTAATCTA CCATCGAGTG GGGAATAACC AACCGAAAGG TTGGCTAATA 120

CCGCGTACGC TTCCGAGTCT TCGGGCTTGG AAGGAAAGCC GCACTGTGAG TGCGGCGCTC 180

TTTGATGAGC TCATATCCTA TCANCTTGTT GGTAGGGTAA CGGCCTACCA AGGCTTTGAC 240

GGGTATCTGG TCTGAGAGGA CGATCAGCCA CACTGGCACT GCGACACGGG CCAGACTCCT 300

ACGGGAGGCA GCAGTAAGGA ATATTGCGCA ATGGGCGAAA CCCNGACGCA GCCACGCCGC 360

GTGGGGGATG AAGGTCTTCG GATTGTAAAC CCCTTTCGGG AGGGAAGATG GAACGAGCAA 420

TCGTTCGGAC GGTACCTCCA GAAGCAGCCA CGGCCAACTT CGTGCCAGCA GCCGCGGTAA 480

TACGAAGGTG GCAAGCGTTG TTCGGATTCA CTGGGCGTAC AGGGTGTGTA GGCGGTTTGG 540

TAAGCCTTCT GTTAAAGCTT CGGGCCCAAC CCGGAAAGCG CAGAGGGTAC TGCCAGGCTA 600

GAGGGTGGGA GAGGAGCGCG GAATTCCCGG TGTAGCGGTG AAATGCGTAG AGATCGGGAG 660

GAAGGCCGGT GGCGAAGGCG GCGCTCTGGA ACATACCTGA CGCTCAGACA CGAAAGCGTG 720

GGGAGCAAAC AGGATTAGAT ACCCTGGTAG TCCACGCCCT AAACTATGGA TACTAAGTGT 780

CGGCGGGTTA CCGCCGGTGC CGCAGCTAAC GCATTAAGTA TCCCGCCTGG GAAGTACGGC 840

CGCAAGGTTG AAACTCAAAG GAATTGACGG GGGCCCGCAC AAGCGGTGGA GCATGTGGTT 900

```
TAATTCGACG CAACGCGAAG AACCTTACCC AGGTTGGACA TGCACGTAGT AGAAAGGTGA    960
AAGCCTGACG AGGTAGCAAT ACCAGCGTGC TCAGGTGCTG CATGGCTGTC GTCAGCTCGT   1020
GCCGTGAGGT GTTGGGTTAA GTCCCGCAAC GAGCGCAACC CCTGCTTTCA GTTGCTGCCG   1080
GGTCATGCCG AACACTCTGA AAGGACTGCC CAGGATAACG GGGAAGGAAG GTGGGGATGA   1140
CGTCAAGTCA GCATGGCCTT TATGCCTGGG GCCACACACG TGCTACAATG GCCGGTACAA   1200
AACGCTGCAA ACCCGTGAGG GGGAGCCAAT CGCAAAAAAC CGGCCTCAGT TCANATTGAG   1260
GTCTGCAACT CGACCTCATG AATGCGGAAT CGCTAGTAAT CGCGGATCAG CACGCCGCGG   1320
TGAATACGTN CCCGGGCCTT GTACACGCCG CCCGTCACAC CACGAAAGCC TGTTGTACCT   1380
GAAGTCGCCC AAGCCAACCG CAAGGAGGCA NGCGCCCACG GTATGGCCGG TGATG        1435
```

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "oligonucleotide primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

```
CGGGAGGGAA GATGGAGC    18
```

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

CCAACCCGGA AAGCGCAGAG 20

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

AGCCTGGCAG TACCCTCT 18

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrococcus mobilis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

CAGCCGGGAG GAAAAGCA 18

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Magnetobacterium bavaricum

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

TGTAGGGAAA GATGATGA 18

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrobacter hamburgensis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

TGTGCGGGAA GATAATGA 18

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospina gracilis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

CGGGTGGGAA GAACAAAA 18

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira marina

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

CATGAGGAAA GATAAAGT 18

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

CGGCAGGGAA GATGGAAC 18

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

CGGGAGGGAA GATGGAGC 18

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

CCGCAGGGAA GATGGAAC 18

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

CGGGAGGGAA GATGGAAC 18

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrobacter

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

CGTGCGGGAA GATAATGA 18

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

CGGCAGGGAA GATGGAAC 18

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira moscoviensis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

CGGGAGGGAA GATGGACG 18

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrococcus mobilis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

TCAACCTGGG AATTGCATCC 20

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Magnetobacterium bavaricum

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

TCAACCCGGG AATTGCCTTG 20

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrobacter hamburgensis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

TCAACTCCAG AACTGCCTTT 20

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospina gracilis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

TCAACCGTGG AATTGCGTTT 20

(2) INFORMATION FOR SEQ ID NO: 33:

(i). SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospina marina

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

TTAACCGGGA AAGGTCGAGA 20

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

CTAACCCGGA AAGTGCGGAG 20

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

CCAACCCGAA AAGCGCAGAG 20

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

CCAACCCGGA AAGCGCAGAG 20

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrobacter

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

TCAACTCCAG AACTGCCTTT 20

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira moscoviensis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

CCAACCCGGA AAGCGCAGAG 20

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrococcus mobilis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

AGCCAAACAG TATCGGAT 18

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Magnetobacterium bavaricum

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

AGTTAAACAG TTTTCAAG 18

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrobacter hamburgensis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

AGACCTTCAG TATCAAAG 18

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospina gracilis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

AGCCGAATAG TTTCAAAC 18

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospina marina

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

AGCTGAATAG TTCCTCTC 18

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

AGCCGAGCAG TCCCCTCC 18

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

AGCCTGGCAG TACCCTCT 18

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

AGCCTGGCAG TACCCCCT 18

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

AGCCTGGCAG TACCGTCT 18

(2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrobacter

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

AGATCCTCAG TATCAAAG 18

(2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira moscoviensis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

AGCCTGGCAG TACCCTCT 18

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Oligonucleotide primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

CCTGTGCTCC ATGCTCCG 18

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrobacter hamburgensis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

CCTGTGCTCC ATGCTCCG 18

(2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospina gracilis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

CCTGTGCAAG GGCCCCGA 18

(2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrococcus mobilis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

CCTGTCATCC GGTTCCCG 18

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira moscoviensis

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

CCTGAGCACG CTGGTATT 18

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospina marina

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

CCTGAGCTCG CTCCCCTT 18

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Magnetobacterium bavaricum

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

CCTGTGCAAG CTCTCCCT 18

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

CCTGAGCAGG ATGGTATT 18

(2) INFORMATION FOR SEQ ID NO: 58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

CCTGAGCACG CTGGTATT 18

(2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Nitrospira

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

CCTGAGCAGG ATGGTGTT 18

CLAIMS

1. A consortium of microorganisms capable of nitrite oxidation in wastewater, which consortium is enriched in members of the *Nitrospira* phylum.

2. An oligonucleotide primer for PCR amplification of *Nitrospira* DNA, said primer comprising at least 12 nucleotides having a sequence selected from the group consisting of:

(i) any one of SEQ ID NO: 1 to SEQ ID NO: 13; and

(ii) a DNA sequence having at least 92% identity with any one of SEQ ID NO: 1 to SEQ ID NO: 13.

3. The oligonucleotide primer of claim 2, wherein said primer has a length of 12 to 50 nucleotides.

4. The oligonucleotide primer of claim 2, wherein said primer has a length of 12 to 22 nucleotides.

5. The oligonucleotide primer of claim 2, wherein said primer sequence is selected from the group consisting of SEQ ID NO: 14, SEQ ID NO: 15 and SEQ ID NO:16.

6. A primer pair for PCR amplification of *Nitrospira* DNA, said primer pair comprising:

(a) a first oligonucleotide of at least 12 nucleotides having a sequence selected from one strand of a bacterial 16S rDNA gene; and

(b) a second oligonucleotide of at least 12 nucleotides having a sequence selected from the other strand of said 16S rDNA gene downstream of said first oligonucleotide sequence; wherein at least one of said first and second oligonucleotides is selected from the group consisting of:

(i) any one of SEQ ID NO: 1 to SEQ ID NO: 13; and

(ii) a DNA sequence having at least 92% identity with any one of SEQ ID NO: 1 to SEQ ID NO: 13.

7. The primer pair of claim 6, wherein said first and second oligonucleotide primers independently have lengths of 12 to 50 nucleotides.

8. The primer pair of claim 6, wherein said first and second oligonucleotide primers independently have lengths of 12 to 22 nucleotides.

9. The primer pair of claim 6, wherein said first oligonucleotide primer sequence is selected from the group consisting of SEQ ID NO: 14 and SEQ ID NO: 15, and said second oligonucleotide primer sequence is SEQ ID NO: 16.

10. A probe for detecting *Nitrospira* DNA, said probe comprising at least 12 nucleotides having a sequence selected from the group consisting of:

(i) any one of SEQ ID NO: 1 to SEQ ID NO: 13; and

(ii) a DNA sequence having at least 92% identity with any one of SEQ ID NO: 1 to SEQ ID NO: 13.

11. The probe of claim 10, wherein said probe has a length of 15 to 50 nucleotides.

12. The probe of claim 10, wherein said probe has a length of 15 to 22 nucleotides.

13. A kit comprising:

at least one primer according to claim 2;

at least one primer pair according to claim 6; or

at least one probe according to claim 10.

14. The kit of claim 13, wherein said kit further includes reagents selected from the group consisting of buffers, salts, detergents, nucleotides and thermostable polymerase.

15. A method of detecting a *Nitrospira* species in a sample, said method comprising the steps of:

(a) lysing cells in said sample to release genomic DNA;

(b) contacting denatured genomic DNA from step (a) with a primer pair according to claim 6;

(c) amplifying *Nitrospira* DNA by cyclically reacting said primer pair with said DNA to produce an amplification product; and

(d) detecting said amplification product.

16. The method according to claim 15, wherein said amplification product has a length of 50 to 1,400 bps.

17. A method of quantitating the level of a *Nitrospira* species in a sample, said method comprising the steps of:

(a) lysing cells in said sample to release genomic DNA;

(b) contacting denatured genomic DNA from step (a) with a primer pair according to claim 6;

(c) amplifying *Nitrospira* DNA by cyclically reacting said primer pair with said DNA to produce an amplification product; and

(d) detecting said amplification product and quantitating the level of said product by comparison with at least one reference standard.

18. The method according to claim 17, wherein said amplification product has a length of 50 to 1,400 bps.

19. A method of detecting a *Nitrospira* species in a sample, said method comprising the steps of:

(a) lysing cells in said sample to release genomic DNA;

(b) contacting denatured genomic DNA from step (a) with a labelled probe according to claim 4 under conditions which allow hybridisation of said genomic DNA said probe;

(c) separating hybridised labeled probe and genomic DNA from unhybridised labeled probe; and

(d) detecting said labeled probe-genomic DNA hybrid.

20. A method of detecting cells of a *Nitrospira* species in a sample, said method comprising the steps of:

(a) treating cells in said sample to fix cellular contents;

(b) contacting said fixed cells from step (a) with a labeled probe according to claim 10 under conditions which allow said probe to hybridise with RNA within said fixed cell;

(c) removing unhybridised probe from said fixed cells; and

(d) detecting said labeled probe-RNA hybrid.

Ridout & Maybee
Suite 2400
One Queen Street East
Toronto, Canada M5C 3B1
Patent Agents of the Applicant

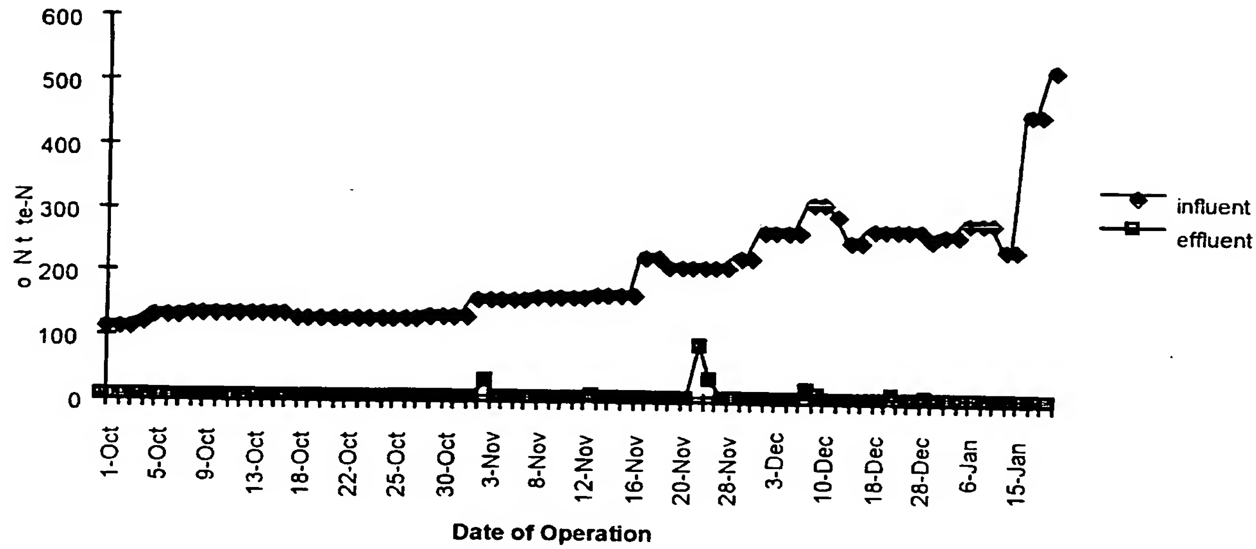

600
500
400
300
200
100
0
o Nt te-N
1-Oct
5-Oct
9-Oct
13-Oct
18-Oct
22-Oct
25-Oct
30-Oct
3-Nov
8-Nov
12-Nov
16-Nov
20-Nov
28-Nov
3-Dec
10-Dec
18-Dec
28-Dec
6-Jan
15-Jan
Date of Operation
influent
effluent

Fig. 1

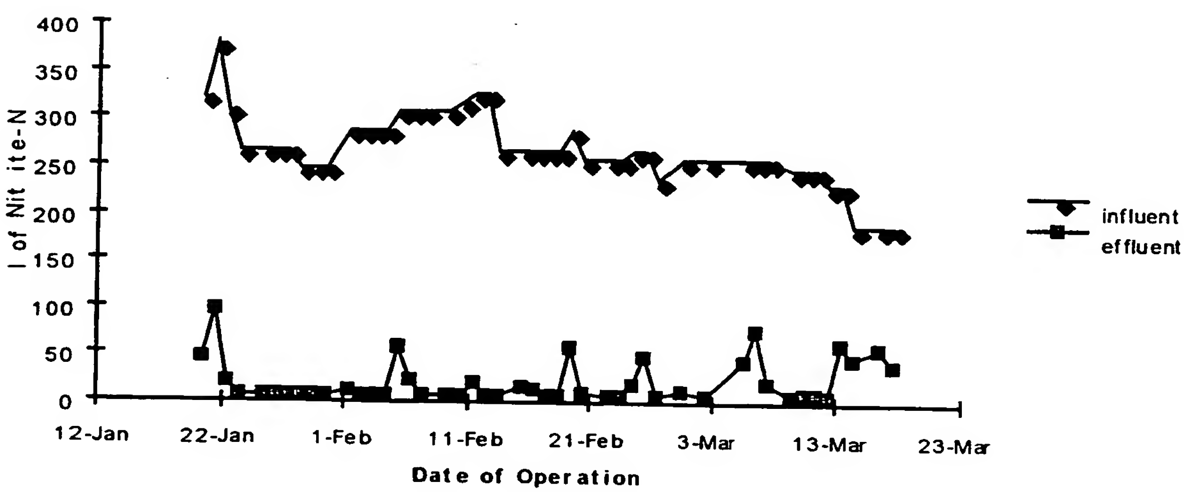
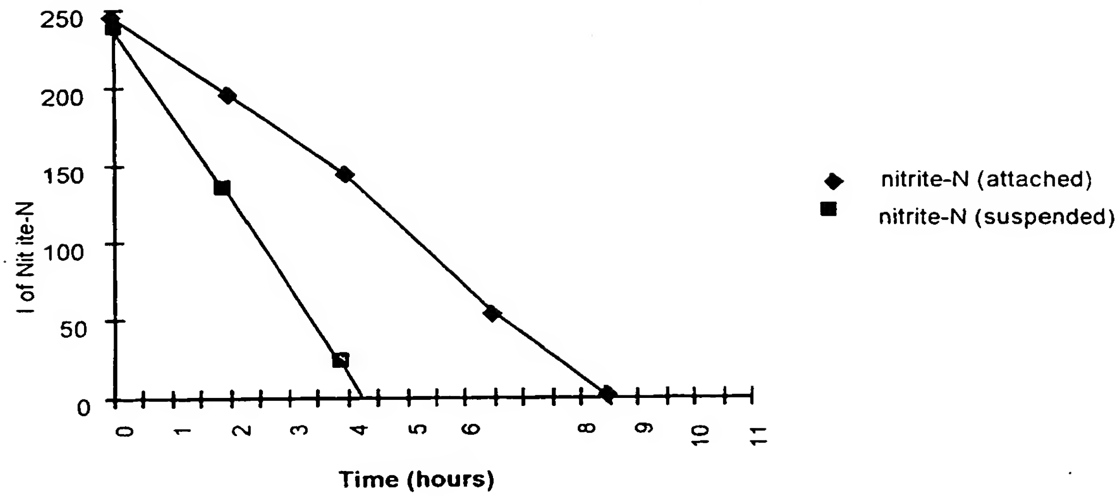
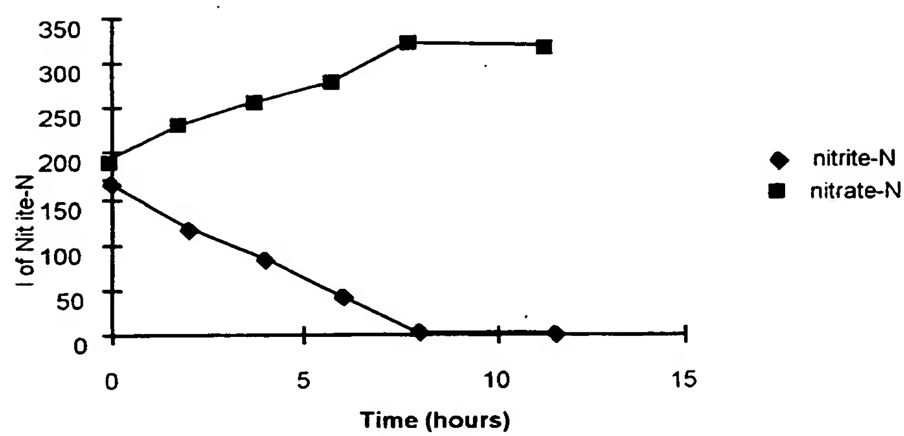

400
350
300
250
200
150
100
50
0
l of Nit ite-N
12-Jan
22-Jan
1-Feb
11-Feb
21-Feb
3-Mar
13-Mar
23-Mar
Date of Operation
influent
effluent

Fig. 2



250
200
150
100
50
0
l of Nit ite-N
0
1
2
3
4
5
6
8
9
10
11
Time (hours)
nitrite-N (attached)
nitrite-N (suspended)

Fig. 3



350
300
250
200
150
100
50
0
l of Nit ite-N
0
5
10
15
Time (hours)
nitrite-N
nitrate-N

Fig. 4

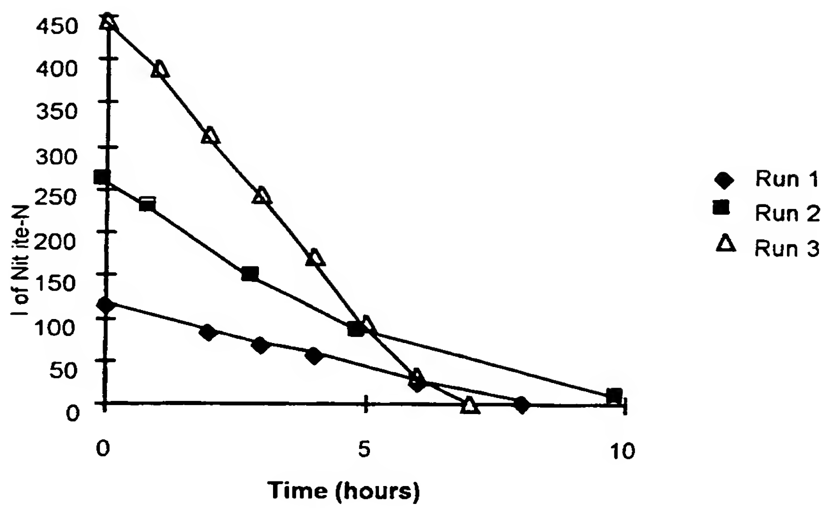

450
400
350
300
250
200
150
100
50
0
l of Nit ite-N
0
5
10
Time (hours)
Run 1
Run 2
Run 3

Fig. 5

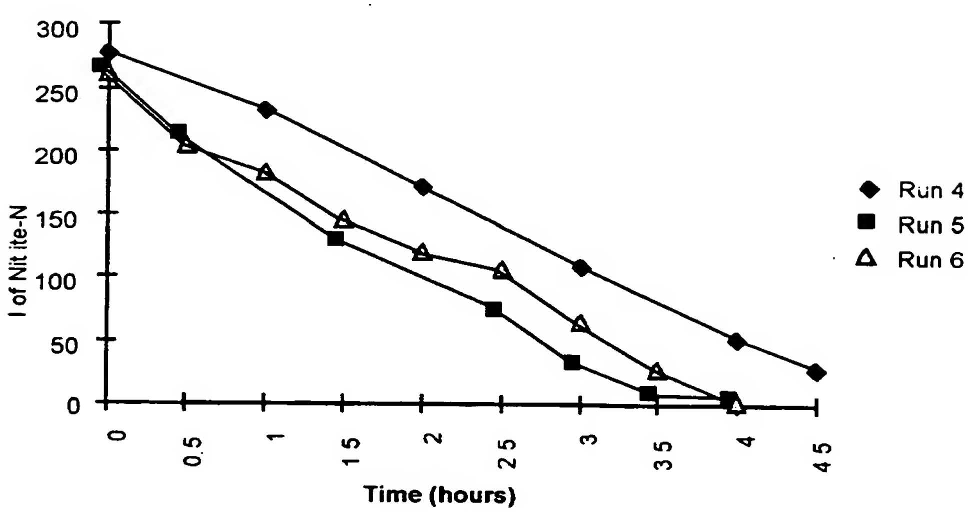
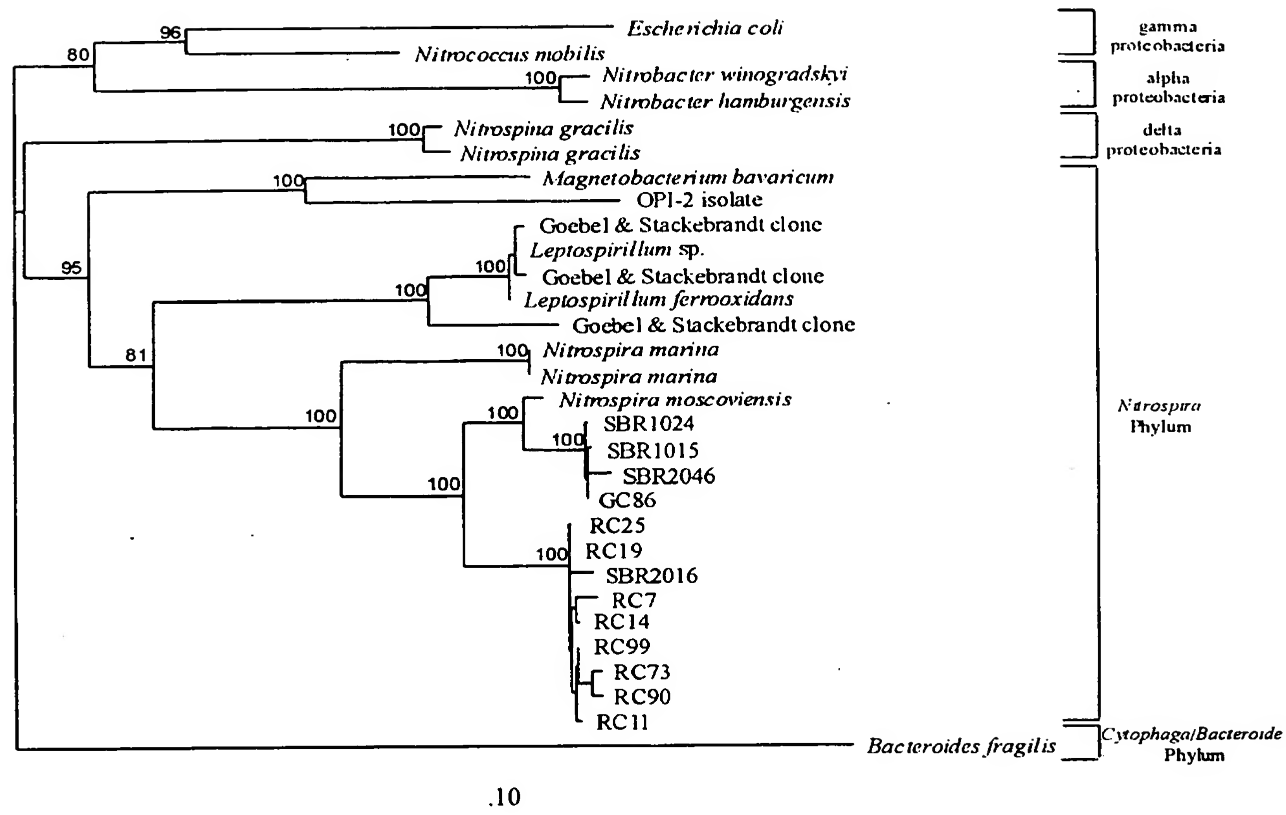

300
250
200
150
100
50
0
l of Nit ite-N
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
Time (hours)
Run 4
Run 5
Run 6

Fig. 6



Escherichia coli
Nitrococcus mobilis
Nitrobacter winogradskyi
Nitrobacter hamburgensis
Nitrospina gracilis
Nitrospina gracilis
Magnetobacterium bavaricum
OPI-2 isolate
Goebel & Stackebrandt clone
Leptospirillum sp.
Goebel & Stackebrandt clone
Leptospirillum ferrooxidans
Goebel & Stackebrandt clone
Nitrospira marina
Nitrospira marina
Nitrospira moscoviensis
SBR1024
SBR1015
SBR2046
GC86
RC25
RC19
SBR2016
RC7
RC14
RC99
RC73
RC90
RC11
Bacteroides fragilis
gamma proteobacteria
alpha proteobacteria
delta proteobacteria
Nitrospira Phylum
Cytophaga/Bacteroide Phylum
96
80
100
100
100
95
100
100
81
100
100
100
100
100
100
.10

Fig. 7

```
[       1                                                      50 ]
SBR1024---------- ---------- ---------- ---------- ----------
SBR1015---------- ---------- ---------- ---------- ----------
 GC86   ---------- ---------- --TCGACCTG CAGGCGGCCG CACTAGTGAT
SBR2046---------- ---------- ---------- ---------- ----------
 RC25   --------GC TCTCCCATAT GGTCGACCTG CAGGCGGCCG CACTAGTGAT
 RC19   ---------- ---------- ---------- ---------- ----------
SBR2016---------- ---------- ---------- ---------- ----------
 RC7    ---------- ---------- ---------- ---------- ----------
 RC14   ---------- ---------- ---------- ---------- ----------
 RC99   ---------- ---------- ---------- ---------- ----------
 RC11   ---------- ---------- ---------- ---------- ----------
 RC73   ---------- ---------- ---------- ---------- ----------
 RC90   ---------- ---------- ---------- ---------- ----------

[      51                                                     100 ]
SBR1024---------- ---------- ---------- ---------- ----------
SBR1015---------- ---------- ---------- ---------- --TAATACAT
 GC86   TAGAGTTTGA TCCTGGCTCA GAACGAACGC TGGCGGCGCG CCTAATACAT
SBR2046---------- ---------- ---------- ---------- ----------
 RC25   TAGAGTTTGA TCCTGGCTCA GAACGAACGC TGGCGGCGCG CCTAATACAT
 RC19   ---------- ---------- ---------- ---------- ----------
SBR2016---------- ---------- ---------- ---------- --TAATACAT
 RC7    ---------- ---------- ---------- ---------- --TAATACAT
 RC14   ---------- ---------- ---------- ---------- ----------
 RC99   ---------- ---------- ---------- ---------- CCTAATACAT
 RC11   ---------- ---------- ---------- ---------- ---AATACAT
 RC73   ---------- ---------- ---------- ---------- ---AATACAT
 RC90   ---------- ---------- ---------- ---------- --TAATACAT

[     101                                                     150 ]
SBR1024-CAAGTCGAG CGAGAAGACG TA........ ...GCAA... ........TA
SBR1015GCAAGTCGAG CGAGAAGACG TA........ ...GCAA... ........TA
 GC86   GCAAGTCGAG CGAGAAGACG TA........ ...GCAA... ........TA
SBR2046---------- CGAGAAGACG TA........ ...GCAA... ........TA
 RC25   GCAAGTCGAG CGAGAAGACG TA........ ...GCAA... ........TA
 RC19   --AAGTCGAG CGAGAAGGTG TA........ ...GCAA... ........TA
SBR2016GCAAGTCGAG CGAGAAGGTG TA........ ...GCAA... ........TA
 RC7    GCAAGTCGAG CGAGAAGGTG TA........ ...GCAA... ........TA
 RC14   ---------- CGAGAAGGTG TA........ ...GCAA... ........TA
 RC99   GCAAGTCGAT CGAGAAGGTG TA........ ...GCAA... ........TA
 RC11   GCAAGTCGAT CGAGAAGGTG TA........ ...GCAA... ........TA
 RC73   GCAAGTCGAT CGANAAGGTG TA........ ...GCAA... ........TA
 RC90   GCAAGTCGAT CGANAAGGTG TA........ ...GCAA... ........TA
```

Fig. 8

```
[    151                                              200 ]
SBR1024CGTTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAACCT
SBR1015CGTTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAGCCT
 GC86   CGTTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAACCT
SBR2046CGTTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAACCT
 RC25   CGTTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAATCT
 RC19   CACTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAATCT
SBR2016CACTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAATCT
 RC7    CACTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAATCT
 RC14   CACTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAATCT
 RC99   CACTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAATCT
 RC11   CACTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAATCT
 RC73   CACTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAATCT
 RC90   CACTTGTAAA GCGGC..... ..GAACGGGT GAGGAATACA TGGGTAATCT

[    201                                              250 ]
SBR1024ACCTTCGAGT GGGGAATAAC TAGCCGAAAG GTTAGCTAAT ACCGCATACG
SBR1015ACCCTCGAGT GGGGAATAAC TAACCGAAAG GTTAGCTAAT ACCGCATACG
 GC86   ACCCTCGAGT GGGGAATAAC TAGCCGAAAG GTTAGCTAAT ACCGCATACG
SBR2046ACCCTCGAGT GGGGAATAAC TAACCGAAAG GTTAGCTAAT ACCGCATACG
 RC25   ACCATCGAGT GGGGAATAAC CAACCGAAAG GTTGGCTAAT ACCGCGTACG
 RC19   ACCATCGAGT GGGGAATAAC CAGCCGAAAG GTTGGCTAAT ACCGCGTACG
SBR2016ACCATCGAGT GGGGAATAAC CAACCGAAAG GTTGGCTAAT ACCGCGTACG
 RC7    ACCATCGAGT GGGGAATAAC CAACCGAAAG GTTGGCTAAT ACCGCGTACG
 RC14   ACCATCGAGT GGGGAATAAC CAACCGAAAG GTTGGCTAAT ACCGCGTACG
 RC99   ACCATCGAGT GGGGAATAAC CAACCGAAAG GTTGGCTAAT ACCGCGTACG
 RC11   ACCATCGAGT GGGGAATAAC CAACCGAAAG GTTGGCTAAT ACCGCGTACG
 RC73   ACCATCGAGT GGGGAATAAC CAACCGAAAG GTTGGCTAAT ACCGCGTACG
 RC90   ACCATCGAGT GGGGAATAAC CAACCGAAAG GTTGGCTAAT ACCGCGTACG

[    251                                              300 ]
SBR1024ACTCCTGGTC .TGC..GGAT CGGGAGAGAA AGCGATACC. ......GTG.
SBR1015GCTCCTGGTC .TGC..GGAT CGGGAGAGAA AGCGATACC. ......GTG.
 GC86   ACTCCTGGTC .TGC..GGAT CGGGAGAGAA AGCGATACC. ......GTG.
SBR2046GCTCCTGGTC .TGC..GGAT CGGGAGAGAA AGCGATACC. ......GTG.
 RC25   CTTCTGAGTC .TTC..GGGT TCGGAAGGAA AGCCGTACT. ......GTG.
 RC19   CTTCCGAGTC .TTC..GGGC TTGGAAGGAA AGCCGCACT. ......GTG.
SBR2016CTTCTGAGCC .TTC..GTGT TCGGAAGGAA AGCCGTACT. ......GTG.
 RC7    CCTCCGAGTC .TTC..GGGT TCGGAGGGAA AGCTGCACT. ......GTG.
 RC14   CCTCCGAGTC .TTC..GGGT TCGGAGGGAA AGCTGCACT. ......GTG.
 RC99   CCTCCGAGTC .TTC..GGGT TCGGAGGGAA AGCTGCACT. ......GTG.
 RC11   CCTCCGAGTC .TTC..GGGT TCGGAGGGAA AGCTGCACT. ......GTG.
 RC73   CCTCCGAGTC .TTC..GGGT TCGGAGGGAA AGCTGCACT. ......GTG.
 RC90   CTTCCGAGTC .TTC..GGGC TTGGAAGGAA AGCCGCACT. ......GTG.
```

Fig. 8 (continued)

```
[    301                                                       350 ]
SBR1024.....GGTAT CGCGCTCTTG GATGGGCTCA TGTCCTATCA GCTTGTTGGT
SBR1015.....GGTAT CGCGCTCTTG GATGGGCTCA TGTCCTATCA GCTTGTTGGT
GC86   .....GGTAT CGCGCTCTTG GATGGGCTCA TGTCCTATCA GCTTGTTGGT
SBR2046.....GGTAT CGCGCTCTTG GATGGGCTCA TGTCCTATCA GCTTGTTGGT
RC25   .....AGTGC GGCGCTCTTT GATGAGCTCA TGTCCTATCA GCTTGTTGGT
RC19   .....AGTGC GGCGCTCTTT GATGAGCTCA TGTCCTATCA GCTTGTTGGT
SBR2016.....AGTGC GGCGCTCTTT GATGAGCTCA TGTCCTATCA GCTTGTTGGT
RC7    .....AGTGT AGCGCTCTTT GATGAGCTCA TGTCCTATCA GCTTGTTGGT
RC14   .....AGTGT AGCGCTCTTT GATGAGCTCA TGTCCTATCA GCTTGTTGGT
RC99   .....AGTGT AGCGCTCTTT GATGAGCTCA TGTCCTATCA GCTTGTTGGT
RC11   .....AGTGT AGCGCTCTTT GATGAGCTCA TGTCCTATCA GCTTGTTGGT
RC73   .....AGTGT AGCGCTCTTT GATGAGCTCA TGTCCTATCA GCTTGTTGGT
RC90   .....AGTGC GGCGCTCTTT GATGAGCTCA TATCCTATCA NCTTGTTGGT

[    351                                                       400 ]
SBR1024GAGGTAACGG CTCACCAAGG CTTCGACGGG TAGCTGGTCT GAGAGGACGA
SBR1015GAGGTAACGG CTCACCAAGG CTTCGACGGG TAGCTGGTCT GAGAGGACGA
GC86   GAGGTAACGG CTCACCAAGG CTTCGACGGG TAGCTGGTCT GAGAGGACGA
SBR2046GAGGTAACGG CTCACCAAGG CTTCGACGGG TAGCTGGTCT GAGAGGACGA
RC25   AGGGTAACGG CCTACCAAGG CTTTGACGGG TAGCTGGTCT GAGAGGACGA
RC19   AGGGTAACGG CCTACCAAGG CTTTGACGGG TAGCTGGTCT GAGAGGACGA
SBR2016AGGGTAACGG CCTACCAAGG CTTTGACGGG TAGCTGGTCT GAGAGGACGA
RC7    AGGGTAACGG CCTACCAAGG CTTTGACGGG TAGCTGGTCT GAGAGGACGA
RC14   AGGGTAACGG CCTACCAAGG CTTTGACGGG TAGCTGGTCT GAGAGGACGA
RC99   AGGGTAACGG CCTACCAAGG CTTTGACGGG TAGCTGGTCT GAGAGGACGA
RC11   AGGGTAACGG CCTACCAAGG CTTTGACGGG TAGCTGGTCT GAGAGGACGA
RC73   AGGGTAACGG CCTACCAAGG CTTTGACGGG TATCTGGTCT GAGAGGACGA
RC90   AGGGTAACGG CCTACCAAGG CTTTGACGGG TATCTGGTCT GAGAGGACGA

[    401                                                       450 ]
SBR1024TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
SBR1015TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
GC86   TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
SBR2046TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
RC25   TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
RC19   TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
SBR2016TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
RC7    TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
RC14   TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
RC99   TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
RC11   TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
RC73   TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
RC90   TCAGCCACAC TGGCACTGCG ACACGGGCCA GACTCCTACG GGAGGCAGCA
```

Fig. 8 (continued)

```
[    451                                                  500 ]
SBR1024GTAAGGAATA TTGCGCAATG GGC.GACAGC CTGACGCAGC NACGCCGCGT
SBR1015GTAAGGAATA TTGCGCAATG GGC.GACAGC CTGACGCAGC NACGCCGCGT
 GC86  GTAAGGAATA TTGCGCAATG GGC.GACAGC CTGACGCAGC NACGCCGCGT
SBR2046GTAAGGAATA TTGCGCAATG GGC.GACAGC CTGACGCAGC GACGCCGCGT
 RC25  GTAAGGAATA TTGCGCAATG GGC.GAAAGC CTGACGCAGC NACGCCGCGT
 RC19  GTAAGGAATA TTGCGCAATG GGC.GAAAGC CTGACGCAGC GACGCCGCGT
SBR2016GTAAGGAATA TTGCGCAATG GGC.GAAAGC CTGACGCAGC NACGCCGCGT
 RC7   GTAAGGAATA TTGCGCAATG GGC.GAAAGC CTGACGCAGC NACGCCGCGT
 RC14  GTAAGGAATA TTGCGCAATG GGC.GAAAGC CTGACGCAGC NACGCCGCGT
 RC99  GTAAGGAATA TTGCGCAATG GGC.GAAAGC CTGACGCAGC CACGCCGCGT
 RC11  GTAAGGAATA TTGCGCAATG GGC.GAAAGC CTGACGCAGC CACGCCGCGT
 RC73  GTAAGGAATA TTGCGCAATG GGC.GAAACC CNGACGCAGC CACGCCGCGT
 RC90  GTAAGGAATA TTGCGCAATG GGC.GAAACC CNGACGCAGC CACGCCGCGT

[    501                                                  550 ]
SBR1024GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGCA GGGAAGATGG
SBR1015GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGCA GGGAAGATGG
 GC86  GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGCA GGGAAGATGG
SBR2046TGGGGATGAA AGTC.TTCCG ATTGTAAACC CCTTTCCGCA GGGAAGATGG
 RC25  GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGGA GGGAAGATGG
 RC19  GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGGA GGGAAGATGG
SBR2016GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGGA GGGAAGATGG
 RC7   GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGGA GGGAAGATGG
 RC14  GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGGA GGGAAGATGG
 RC99  GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGGA GGGAAGATGG
 RC11  GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGGA GGGAAGATGG
 RC73  GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGGA GGGAAGATGG
 RC90  GGGGGATGAA GGTC.TTCGG ATTGTAAACC CCTTTCGGGA GGGAAGATGG

[    551                                                  600 ]
SBR1024AACGG..... .GTAA..... ...CCGTTCG GACGGTACCT GCAGAAGCAG
SBR1015AACGG..... .GTAA..... ...CCGTTCG GACGGTACCT GCAGAAGCAG
 GC86  AACGG..... .GTAA..... ...CCGTTCG GACGGTACCT GCAGAAGCAG
SBR2046AACGG..... .GTAA..... ...CCGTTCG GACGGTACCT GCAGAAGCAG
 RC25  AGCGA..... .GCAA..... ...TCGTTCG GACGGTACCT CCAGAAGCAG
 RC19  AGCCA..... .GCAA..... ...TCGTTCG GACGGTACCT CCAGAAGCAG
SBR2016AGCGA..... .GCAA..... ...TCGTTCG GACGGTACCT CCAGAAGCAG
 RC7   AGCGA..... .GCAA..... ...TCGTTCG GACGGTACCT CCAGAAGCAG
 RC14  AGCGA..... .GCAA..... ...TCGTTCG GACGGTACCT CCAGAAGCAG
 RC99  AGCGA..... .GCAA..... ...TCGTTCG GACGGTACCT CCAGAAGCAG
 RC11  AGCGA..... .GCAA..... ...TCGTTCG GACGGTACCT CCAGAAGCAG
 RC73  AACGA..... .GCAA..... ...TCGTTCG GACGGTACCT CCAGAAGCAG
 RC90  AACGA..... .GCAA..... ...TCGTTCG GACGGTACCT CCAGAAGCAG
```

Fig. 8 (continued)

[601					650]
SBR1024	CCACGGCTAA	CTTCGTGCCA	GCAGCCGCGG	TAATACGAAG	GTGGCAAGCG
SBR1015	CCACGGCTAA	CTTCGTGCCA	GCAGCCGCGG	TAATACGAAG	GTGGCAAGCG
GC86	CCACGGCTAA	CTTCGTGCCA	GCAGCCGCGG	TAATACGAAG	GTGGCAAGCG
SBR2046	CCACGGCTAA	CTTCGTGCCA	GCAGCCGCGG	TAATACGAAG	GTGGCAAGCG
RC25	CCACGGCCAA	CTTCGTGCCA	GCAGCCGCGG	TAATACGAAG	GTGGCAAGCG
RC19	CCACGGCCAA	CTTCGTGCCA	GCAGCCGCGG	TAATACGAAG	GTGGCAAGCG
SBR2016	CCACGGCCAA	CTTCGTGCCA	GCAGCCGCGG	TAATACGAAG	GTGGCAAGCG
RC7	CCACGGCCAA	CTTCGTGCCA	GCAGCCGCGG	TAATACGAAG	GTGGCAAGCG
RC14	CCACGGCCAA	CTTCGTGCCA	GCAGCCGCGG	TAATACGAAG	GTGGCAAGCG
RC99	CCACGGCCAA	CTTCGTGCCA	GCAGCCGCGG	TAATACGAAG	GTGGCAAGCG
RC11	CCACGGCCAA	CTTCGTGCCA	GCAGCCGCGG	TAATACGAAG	GTGGCAAGCG
RC73	CCACGGCCAA	CTTCGTGCCA	GCAGCCGCGG	TAATACGAAG	GTGGCAAGCG
RC90	CCACGGCCAA	CTTCGTGCCA	GCAGCCGCGG	TAATACGAAG	GTGGCAAGCG

[651					700]
SBR1024	TTGTTCGGAT	TTACTGGGCG	TACAGGGAGC	GTAGGCGGTT	GGGTAAGCCC
SBR1015	TTGTTCGGAT	TTACTGGGCG	TACAGGGAGC	GTAGGCGGTT	GGGTAAGCCC
GC86	TTGTTCGGAT	TTACTGGGCG	TACAGGGAGC	GTAGGCGGTT	GGGTAAGCCC
SBR2046	TTGTTCGGAT	TTACTGGGCG	TACAGGGAGC	GTAGGCGGTT	GGGTAAGCCC
RC25	TTGTTCGGAT	TCACTGGGCG	TACAGGGTGT	GTAGGCGGTT	TGGTAAGCCT
RC19	TTGTTCGGAT	TCACTGGGCG	TACAGGGTGT	GTANGCGGTT	TGGTAAGCCT
SBR2016	TTGCTTGGAT	TCACTGGGCG	TACAGGGTGT	GTAGGCGGTT	TGGTAAGCCT
RC7	TTGTTCGGAT	TCACTGGGCG	TACAGGGTGT	GTAGGCGGTT	TGGTAAGCCT
RC14	TTGTTCGGAT	TCACTGGGCG	TACAGGGTGT	GTAGGCGGTT	TGGTAAGCCT
RC99	TTGTTCGGAT	TCACTGGGCG	TACAGGGTGT	GTAGGCGGTT	TGGTAAGCCT
RC11	TTGTTCGGAT	TCACTGGGCG	TACAGGGTGT	GTAGGCGGTT	TGGTAAGCCT
RC73	TTGTTCGGAT	TCACTGGGCG	TACAGGGTGT	GTAGGCGGTT	TGGTAAGCCT
RC90	TTGTTCGGAT	TCACTGGGCG	TACAGGGTGT	GTAGGCGGTT	TGGTAAGCCT

[701					750]
SBR1024	TCCGTGAAAT	CTCCGGGCCT	AACCCGGAAA	GTGCGGAGGG	GACTGCTCGG
SBR1015	TCCGTGAAAT	CTCCGGGCCT	AACCCGGAAA	GTGCGGAGGG	GACTGCTCGG
GC86	TCCGTGAAAT	CTCCGGGCCT	AACCCGGAAA	GTGCGGAGGG	GACTGCTCGG
SBR2046	TCCGTGAAAT	CTCCGGGCCT	AACCCGGAAA	GTGCGGAGGG	GACTGCTCGG
RC25	TCTGTTAAAG	CTTCGGGCCC	AACCCGGAAA	GCGCAGACGG	TACTGCCAGG
RC19	TCTGTTAAAG	CTTCGGGCCC	AACCCGGAAA	GCGCAGAGGG	TACTGCCAGG
SBR2016	TCTGTTAAAG	CTTCGGGCCC	AACCCGAAAA	GCGCAGAGGG	TACTGCCAGG
RC7	TCTGTTAAAG	CTTCGGGCCC	AACCCGGAAA	GCGCAGGGGG	TACTGCCAGG
RC14	TCTGTTAAAG	CTTCGGGCCC	AACCCGGAAA	GCGCAGAGGG	TACTGCCAGG
RC99	TCTGTTAAAG	CTTCGGGCCC	AACCCGGAAA	GCGCAGAGGG	TACTGCCAGG
RC11	TCTGTTAAAG	CTTCGGGCCC	AACCCGGAAA	GCGCAGAGGG	TACTGCCAGG
RC73	TCTGTTAAAG	CTTCGGGCCC	AACCCGGAAA	GCGCAGAGGG	TACTGCCAGG
RC90	TCTGTTAAAG	CTTCGGGCCC	AACCCGGAAA	GCGCAGAGGG	TACTGCCAGG

Fig. 8 (continued)

[751 800]

SBR1024	CTAGAGGATG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
SBR1015	CTAGAGGATG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
GC86	CTAGAGGATG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
SBR2046	CTAGAGGATG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
RC25	CTAGAGGGTG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
RC19	CTAGAGGGTG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
SBR2016	CTAGAGGGTG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
RC7	CTAGAGGGTG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
RC14	CTAGAGGGTG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
RC99	CTAGAGGGTG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
RC11	CTAGAGGGTG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
RC73	CTAGAGGGTG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG
RC90	CTAGAGGGTG	GGAGAGGAGC	GCGGAATTCC	CGGTGTAGCG	GTGAAATGCG

[801 850]

SBR1024	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATTTC
SBR1015	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATTTC
GC86	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATTTC
SBR2046	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATTTC
RC25	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATACC
RC19	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATGCC
SBR2016	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATACC
RC7	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATACC
RC14	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATACC
RC99	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATACC
RC11	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATACC
RC73	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATACC
RC90	TAGAGATCGG	GAGGAAGGCC	GGTGGCGAAG	GCGGCGCTCT	GGAACATACC

[851 900]

SBR1024	TGACGCTGAG	GCTCGAAAGC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG
SBR1015	TGACGCTGAG	GCTCGAAAGC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG
GC86	TGACGCTGAG	GCTCGAAAGC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG
SBR2046	TGACGCTGAG	GCTCGAAAGC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG
RC25	TGACGCTGAG	ACACGAAAGC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG
RC19	TGACGCTGAG	ACACGAAAGC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG
SBR2016	TGACGCTGAG	ACACGAAAAC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG
RC7	TGACGCTGAG	ACACGAAAGC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG
RC14	TGACGCTGAG	ACACGAAAGC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG
RC99	TGACGCTGAG	ACACGAAAGC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG
RC11	TGACGCTGAG	ACACGAAAGC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG
RC73	TGACGCTGAG	ACACGAAAGC	GTGGGGNGCA	AACAGGATTA	GATACCCTGG
RC90	TGACGCTCAG	ACACGAAAGC	GTGGGGAGCA	AACAGGATTA	GATACCCTGG

Fig. 8 (continued)

```
[    901                                              950 ]
SBR1024TAGTCCACGC CTTAAACGAT GGATACTAAG TGTCGGCGG. ..........
SBR1015TAGTCCACGC CTTAAACGAT GGATACTAAG TGTCGGCGG. ..........
 GC86  TAGTCCACGC CTTAAACGAT GGATACTAAG TGTCGGCGG. ..........
SBR2046TAGTCCACGC CTTAAACGAT GGATACTAAG TGTCGGCGG. ..........
 RC25  TAGTCCACGC CCTAAACTAT GGATACTAAG TGTCGGCGG. ..........
 RC19  TAGTCCACGC CCTAAACTAT GGATACTAAG TGTCGGCGG. ..........
SBR2016TAGTCCACGC CCTAAACTAT GGATACTAAG TGTCGGCGG. ..........
 RC7   TAGTCCACGC CCTAAGCTAT GGATACTAAG TGTCGGCGG. ..........
 RC14  TAGTCCACGC CCTAAACTAT GGATACTAAG TGTCGGCGG. ..........
 RC99  TAGTCCACGC CCTAAACTAT GGATACTAAG TGTCGGCGG. ..........
 RC11  TAGTCCACGC CCTAAACTAT GGATACTAAG TGTCGGCGG. ..........
 RC73  TAGTCCACGC CCTAAACTAT GGATACTAAG TGTCGGCGG. ..........
 RC90  TAGTCCACGC CCTAAACTAT GGATACTAAG TGTCGGCGG. ..........

[    951                                             1000 ]
SBR1024.........G TTA....... .......... .CCGCCGGTG CCGCAGCTAA
SBR1015.........G TTA....... .......... .CCGCCGGTG CCGCAGCTAA
 GC86  .........G TTA....... .......... .CCGCCGGTG CCGCAGCTAA
SBR2046.........G TTA....... .......... .CCGCCGGTG CCGCAGCTAA
 RC25  .........G TTA....... .......... .CCGCCGGTG CCGCAGCTAA
 RC19  .........G TTA....... .......... .CCGCCGGTG CCGCAGCTAA
SBR2016.........G TTA....... .......... .CCGCCGGTG CCGCAGCTAA
 RC7   .........G TTA....... .......... .CCGCCGGTG CCGCAGCCAA
 RC14  .........G TTA....... .......... .CCGCCGGTG CCGCAGCTAA
 RC99  .........G TTA....... .......... .CCGCCGGTG CCGCAGCTAA
 RC11  .........G TTA....... .......... .CCGCCGGTG CCGCAGCTAA
 RC73  .........G TTA....... .......... .CCGCCGGTG CCGCAGCTAA
 RC90  .........G TTA....... .......... .CCGCCGGTG CCGCAGCTAA

[   1001                                             1050 ]
SBR1024CGCATTAAGT ATCCCGCCTG GGAAGTACGG CCGCAAGGTT GAAACTCAAA
SBR1015CGCATTAAGT ATCCCGCCTG GGAAGTACGG CCGCAAGGTT GAAACTCAAA
 GC86  CGCATTAAGT ATCCCGCCTG GGAAGTACGG CCGCAAGGTT GAAACTCAAA
SBR2046CGCATTAAGT ATCCCGCCTG GGAAGTACGG CCGCAAGGTT GAAACTCAAA
 RC25  CGCATTAAGT ATCCCGCCTG GGAAGTACGG CCGCAAGGTT GAAACTCAAA
 RC19  CGCATTAAGT ATCCCGCCTG GGAAGTACGG CCGCAAGGTT GAAACTCAAA
SBR2016CGCATTAAGT ATCCCGCCTG GGAGGTACGG CCGCAAGGTT GAAACTCAAA
 RC7   CGCGTTAAGT ATCCCGCCTG GGAAGTACGG CCGCAAGGTT GAAACTCAAA
 RC14  CGCATTAAGT ATCCCGCCTG GGAAGTACGG CCGCAAGGTT GAAACTCAAA
 RC99  CGCATTAAGT ATCCCGCCTG GGAAGTACGG CCGCAAGGTT GAAACTCAAA
 RC11  CGCATTAAGT ATCCCGCCTG GGAAGTACGG CCGCAAGGTT GAAACTCAAA
 RC73  CGCATTAAGT ATCCCGCCTG GGAAGTACGG CCGCAAGGTT GAAACTCAAA
 RC90  CGCATTAAGT ATCCCGCCTG GGAAGTACGG CCGCAAGGTT GAAACTCAAA
```

Fig. 8 (continued)

```
[  1051                                                      1100 ]
SBR1024GGAATTGACG GGGGCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC
SBR1015GGAATTGACG GGGGCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC
 GC86  GGAATTGACG GGGGCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC
SBR2046GGAATTGACG GGGCCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC
 RC25  GGAATTGACG GGGGCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC
 RC19  GGAATTGACG GGGGCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC
SBR2016GGAATTGACG GGGGCCCGCA CAAGCGGTGG AGCTTGTGGT TTAATTCGAC
 RC7   GGAATTGACG GGGGCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC
 RC14  GGAATTGACG GGGGCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC
 RC99  GGAATTGACG GGGGCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC
 RC11  GGAATTGACG GGGGCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC
 RC73  GGGATTGACG GGGGCCCGCA CAAGCGGTGG GGCATGTGGT TTAATTCGAC
 RC90  GGAATTGACG GGGGCCCGCA CAAGCGGTGG AGCATGTGGT TTAATTCGAC

[  1101                                                      1150 ]
SBR1024GCAACGCGAA GAACCTTA.C CCAGGCTGGA CATG...... ...CAGGTAG
SBR1015GCAACGCGAA GAACCTTA.C CCAGGCTGGA CATG...... ...CAGGTAG
 GC86  GCAACGCGAA GAACCTTA.C CCAGGCTGGA CATG...... ...CAGGTAG
SBR2046GCAACGCGAA GAACCTTA.C CCAGGCAGGA CATG...... ...CAGGTAG
 RC25  GCAACGCGAA GAACCTTA.C CCAGGTTGGA CATG...... ...CACGTAG
 RC19  GCAACGCGAA GAACCTTA.C CCAGGTTGGA CATG...... ...CACGTAG
SBR2016GCAACGCGAA GAACCTTA.C CCAGGTTGGA CATG...... ...CACGTAG
 RC7   GCAACGCGAA GAACCTTA.C CCAGGTTGGA CATG...... ...CACGTAG
 RC14  GCAACGCGAA GAACCTTA.C CCAGGTTGGA CATG...... ...CACGTAG
 RC99  GCAACGCGAA GAACCTTA.C CCAGGTTGGA CATG...... ...CACGTAG
 RC11  GCAACGCGAA GAACCTTA.C CCAGGTTGGA CATG...... ...CACGTAG
 RC73  GCAACGCGAA GAACCTTA.C CCAGGTTGGA CATG...... ...CACGTAG
 RC90  GCAACGCGAA GAACCTTA.C CCAGGTTGGA CATG...... ...CACGTAG

[  1151                                                      1200 ]
SBR1024TAGAAGGGT. .GAAA..GCC TAACGAGGTA .....GCAA. ....TACCAT
SBR1015TAGAAGGGT. .GAAA..GCC TAACGAGGTA .....GCAA. ....TACCAT
 GC86  TAGAAGGGT. .GAAA..GCC TAACGAGGTA .....GCAA. ....CACCAT
SBR2046TAGAAGGGT. .GAAA..GCC TAACGAGGTA .....GCAA. ....TACCAT
 RC25  TAGAAAGGT. .GAAA..GCC TGACGAGGTA .....GCAA. ....TACCAG
 RC19  TAGAAAGGT. .GAAA..GNC TAACGAGGTA .....GCAA. ....TACCAG
SBR2016TAGAAAGGT. .GAAA..GCC TGACGAGGTA .....GCAA. ....TACCAG
 RC7   TAGAAAGGT. .GAAA..GCC TGACGAGGTA .....GCAA. ....TACCAG
 RC14  TAGAAAGGT. .GAAA..GCC TGACGAGGTA .....GCAA. ....TACCAG
 RC99  TAGAAAGGT. .GAAA..GCC TGACGAGGTA .....GCAA. ....TACCAG
 RC11  TANAAAGGT. .GAAA..GCC TGACGAGGTA .....GCAA. ....TACCAG
 RC73  TNGAAAGGT. .GAAA..GCC TGACGAGGTA .....GCAA. ....TACCAG
 RC90  TAGAAAGGT. .GAAA..GCC TGACGAGGTA .....GCAA. ....TACCAG
```

Fig. 8 (continued)

```
[   1201                                                    1250 ]
SBR1024CCTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG
SBR1015CCTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG
 GC86  CCTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG
SBR2046CCTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG
 RC25  CGTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG
 RC19  CGTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG
SBR2016CGTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG
 RC7   CGTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG
 RC14  CGTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG
 RC99  CGTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG
 RC11  CGTGCTCAGG TGCTGCATGG CTGTCTTCAG CTCGTGCCGT GAGGTGTTGG
 RC73  CGTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG
 RC90  CGTGCTCAGG TGCTGCATGG CTGTCGTCAG CTCGTGCCGT GAGGTGTTGG

[   1251                                                    1300 ]
SBR1024GTTAAGTCCC GCAACGAGCG CAACCCCTGT CTTCAGTTAC CAACGG....
SBR1015GTTAAGTCCC GCAACGAGCG CAACCCCTGT CTTCAGTTAC CAACGG....
 GC86  GTTAAGTCCC GCAACGAGCG CAACCCCTGT CTTCAGTTAC CAACGG....
SBR2046GTTAAGTCCC GCAACGAGCG CAACCCCTGT CTTCAGTTAC CAACGG....
 RC25  GTTAAGTCCC GCAACGAGCG CAACCCCTGC TTTCAGTTGC TACCGG....
 RC19  GTTAAGTCCC GCAACGAGCG CAACCCCTGC TTTCAGTTGC TACCGG....
SBR2016GTTAAGTCCC GCAACGAGCG CAACCCCTGC TTTCAGTTGC TACCGG....
 RC7   GTTAAGTCCC GCAACGAGCG CAACCCCTGC TTTCAGTTGC TACCGG....
 RC14  GTTAAGTCCC GCAACGAGCG CAACCCCTGC TTTCAGTTGC TACCGG....
 RC99  GTTAAGTCCC GCAACGAGCG CAACCCCTGC TTTCAGTTGC TACCGG....
 RC11  GTTAAGTCCC GCAACGAGCG CAACCCCTGC TTTCAGTTGC TACCGG....
 RC73  GTTAAGTCCC GCAACGAGCG CAACCCCTGC TTTCAGTTGC TACCGG....
 RC90  GTTAAGTCCC GCAACGAGCG CAACCCCTGC TTTCAGTTGC TGCCGG....

[   1301                                                    1350 ]
SBR1024GTCATG.... CCGGGAACTC TGGAGAGACT GCCCAGGAGA ACGGG.GAGG
SBR1015GTCATG.... CCGGGAACTC TGGAGAGACT GCCCAGGAGA ACGGGGGAGG
 GC86  GTCATG.... CCGGGAACTC TGGAGAGACT GCCCAGGAGA ACGGG.GAGG
SBR2046GTCATG.... CCGGGAACTC TGGAGAGACT GCCCAGGAGA ACGGG.GAGG
 RC25  GTCATG.... CCGAGCACTC TGAAAGGACT GCCCAGGATA ACGGG.GAGG
 RC19  GTCATG.... CCGAGCACTC TGAAAGGACT GCCCAGGATA ACGGG.GAGG
SBR2016GTCATG.... CCGAGCACTC TGAAAGGACT GCCCAGGATA ACGGG.GAGG
 RC7   GTCATG.... CCGAGCACTC TGAAAGGACT GCCCAGGATA ACGGGGGAGG
 RC14  GTCATG.... CCGAGCACTC TGAAAGGACT GCCCAGGATA ACGGG.GAGG
 RC99  GTCATG.... CCGAGCACTC TGAAAGGACT GCCCAGGATA ACGGGGAAGG
 RC11  GTCATG.... CCGAACACTC TGAAAGGACT GCCCAGGATA ACGGGGAAGG
 RC73  GTCATG.... CCGAACACTC TGAAAGGACT GCCCAGGATA ACGGGGAAGG
 RC90  GTCATG.... CCGAACACTC TGAAAGGACT GCCCAGGATA ACGGGGAAGG
```

Fig. 8 (continued)

[1351					1400]
SBR1024	AAGGTGGGGA	TGACGTCAAG	TCAGCATGGC	CTTTATGCCT	GGGGCCACAC
SBR1015	AAGGTGGGGA	TGACGTCAAG	TCAGCATGGC	CTTTATGCCT	GGGGCCACAC
GC86	AAGGTGGGGA	TGACGTCAAG	TCAGCATGGC	CTTTATGCCT	GGGGCCACAC
SBR2046	AAGGTGGGGA	TGACGTCAAG	TCAGCATGGC	CTTTATGCCT	GGGGCCACAC
RC25	AAGGTGGGGA	TGACGTCAAG	TCAGCATGGC	CTTTATGCCT	GGGGCCACAC
RC19	AAGGTGGGGA	TGACGTCAAG	TCAGCATGGC	CTTTATGCCT	GGGGCCACAC
SBR2016	AAGGTGGGGA	TGACGTCAAG	TCAGCATGGC	CTTTATGCCT	GGGGCCACAC
RC7	AAGGTGGGGA	TGACGTCAAG	TCAGCATGGC	CTTTATGCCT	GGGGCCACAC
RC14	AAGGTGGGGA	TGACGTCAAG	TCAGCATGGC	CTTTATGCCT	GGGGCCACAC
RC99	AAGGTGGGGA	TGACGTCAAG	TCAGCATGGC	CTTTATGCCT	GGGGCCACAC
RC11	AAGGTGGGGA	TGACGTCAAG	TCAGCATGGC	CTTTATGCCT	GGGGCCACAC
RC73	AAGGTGGGGA	TGACGTCAAG	TCAGCATGGC	CTTTATACCT	GGGGCCACAC
RC90	AAGGTGGGGA	TGACGTCAAG	TCAGCATGGC	CTTTATGCCT	GGGGCCACAC

[1401					1450]
SBR1024	ACGTGCTACA	ATGGCCGGTA	CAAAGCGCTG	CAAACCC.GT	AAGGGGGAGC
SBR1015	ACGTGCTACA	ATGGCCGGTA	CAAAGCGCTG	CAAACCC.GT	AAGGGGGAGC
GC86	ACGTGCTACA	ATGGCCGGTA	CAAAGCGCTG	CAAACCC.GT	AAGGGGGAGC
SBR2046	ACGTGCTACA	ATGGCCGGTA	CAAAGCGCTG	CAAACCC.GT	AAGGGGGAGC
RC25	ACGTGCTACA	ATGGCCGGTA	CAAAGCGCTG	CAAACCC.GT	GAGGGGGAGC
RC19	ACGTGCTACA	ATGGCCGGTA	CAAAGCGCTG	CAAACCC.GT	GAGGGGGAGC
SBR2016	ACGTGCTACA	ATGGCCGGTA	CAAAGCGCTG	CAAACCC.GT	GAGGGGGAGC
RC7	ACGTGCTACA	ATGGCCGGTA	CAAAACGCTG	CAAACCC.GT	GAGGGGGAGC
RC14	ACGTGCTACA	ATGGCCGGTA	TAAAACGCTG	CAAACCC.GT	GAGGGGGAGC
RC99	ACGTGCTACA	ATGGCCGGTA	CAAAACGCTG	CAAACCC.GT	GAGGGGGAGC
RC11	ACGTGCTACA	ATGGCCGGTA	CAAAGCGCTG	CAAACCC.GT	GAGGGGGAGC
RC73	ACGTGCTACA	ATGGCCGGTA	CAAAACGCTG	CAAACCC.GT	GAGGGGGAGC
RC90	ACGTGCTACA	ATGGCCGGTA	CAAAACGCTG	CAAACCC.GT	GAGGGGGAGC

[1451					1500]
SBR1024	CAATCCCAAA	AAACCGGCCT	CAGTTCAGAT	TGAGGTCTGC	AACTCGACCT
SBR1015	CAATCGCAAA	AAACCGGCCT	CAGTTCAGAT	TGAGGTCTGC	AACTCGACCT
GC86	CAATCGCAAA	AAACCGGCCT	CAGTTCAGAT	TGAGGTCTGC	AACTCGACCT
SBR2046	CAATCGCAAA	AAACCGGCCT	CAGTTCAGAT	TGAGGTCTGC	AACTCGACCT
RC25	CAATCGCAAA	AAACCGGCCT	CAGTTCAGAT	TGAGGTCTGC	AACTCGACCT
RC19	CAATCGCAAA	AAACCGGCCT	CAGTTCAGAT	TGAGGTCTGC	AACTCGACCT
SBR2016	CAATCGCAAA	AAACCGGCCT	CAGTTCAGAT	TGAGGTCTGC	AACTCGACCT
RC7	CAATCGCAAA	AAACCGGCCT	CAGTTCAGAT	TGAGGTCTGC	AACTCGACCT
RC14	CAATCGCAAA	AAACCGGCCT	CAGTTCAGAT	TGAGGTCTGC	AACTCGACCT
RC99	CAATCGCAAA	AAACCGGCCT	CAGTTCAGAT	TGAGGTCTGC	AACTCGACCT
RC11	CAATCGCAAA	AAACCGGCCT	CAGTTCAGAT	TGAGGTCTGC	AACTCGACCT
RC73	CAATCGCAAA	AAACCGGCCT	CAGTTCAGAT	TGAGGTCTGC	AACTCGACCT
RC90	CAATCGCAAA	AAACCGGCCT	CAGTTCANAT	TGAGGTCTGC	AACTCGACCT

Fig. 8 (continued)

```
[   1501                                                        1550 ]
SBR1024CATGAAGGCG GAATCGCTAG TAATCCCGGA TCAG.CACGC CGGGGTGAAT
SBR1015CATGAAGGCG GAATCGCTAG TAATCCCGGA TCAG.CACGC CGGGGTGAAT
 GC86   CATGAAGGCG GAATCGCTAG TAATCCCGGA TCAG.CACGC CGGGGTGAAT
SBR2046CATGAAGGCG GAATCGCTAG TAATCCCGGA TCAG.CACGC CGGGGTGAAT
 RC25   CATGAAGGCG GAATCGCTAG TAATCGCGGA TCAG.CACGC CGCGGTGAAT
 RC19   CATGAAGGCG GAATCGCTAG TAATCGCGGA TCAG.CACGC CGCGGTGAAT
SBR2016CATGAAGGCG GAATCGCTAG TAATCGCGGA TCAG.CACGC CGCGGTGAAT
 RC7    CATGAAGGCG GAATCGCTAG TAATCGCGGA TCAG.CACGC CGCGGTGAAT
 RC14   CATGAAGGCG GAATCGCTAG TAATCGCGGA TCAG.CACGC CGCGGTGAAT
 RC99   CATGAAGGCG GAATCGCTAG TAATCGCGGA TCAG.CACGC CGCGGTGAAT
 RC11   CATGAAGGCG GAATCGCTAG TAATCGCGGA TCAG.CACGC CGCGGTGAAT
 RC73   CATGAATGCG GAATCGCTAG TAATCGCGGA TCAG.CACGC CGCGGTGAAT
 RC90   CATGAATGCG GAATCGCTAG TAATCGCGGA TCAG.CACGC CGCGGTGAAT

[   1551                                                        1600 ]
SBR1024ACGTTCCCGG GCCTTGTACA CACCGCCCGT CACACCACGA AAGTTTGTTG
SBR1015ACGTTCCCGG ACCTTGTACA CACCGCCCGT CACACCACGA AAGTTTGTTG
 GC86   ACGTTCCCGG GCCTTGTACA CACCGCCCGT CACACCACGA AAGTTTGTTG
SBR2046ACGTTCCCGG GCCTTGTACA CACCGCCCGT CACACCACGA AAGTTTGTTG
 RC25   ACGTTCCCGG GCCTTGTACA CACCGCCCGT CACACCACGA AAGCCTGTTG
 RC19   ACGTTCCCGG GCCTTGTACA CACCGCCCGT CACACCACGA AAGCCTGTTG
SBR2016ACGTTCCCGG GCCTTGTACA CACCGCCCGT CACACCACGA AAGCCTGTTG
 RC7    ACGTTCCCGG GCCTTGTGCA CACCGCCCGT CACACCACGA AAGCCTGTTG
 RC14   ACGTTCCCGG GCCTTGTACA CACCGCCCGT CACACCACGA AAGCCTGTTG
 RC99   ACGTNCCCGG GCCTTGTACA CACCGCCCGT CACACCACGA AAGCCTGTTG
 RC11   ACGTNCCCGG GCCTTGTACA CACCGCCCGT CACACCACGA AAGCCTGTTG
 RC73   ACGTNCCCGG GCCTTGTACA CACCGCCCGT CACACCACGA AAGCCTGTTG
 RC90   ACGTNCCCGG GCCTTGTACA CGCCGCCCGT CACACCACGA AAGCCTGTTG

[   1601                                                        1650 ]
SBR1024TACCTGAAGT CGTTGGCGCC AACC...... GCAA...... GGAGGCAGAC
SBR1015TACCTGAAGT CGTTGGCGCC AACC...... GCAA...... GGAG------
 GC86   TACCTGAAGT CGTTGGCGCC AACC...... GCAA...... GGGGGCAGAC
SBR2046TACCTGAAGT CGTTGGCGCC AACC...... GCAA...... GGAGGCAGAC
 RC25   TACCTGAAGT CGCCCAAGCC AACC...... GCAA...... GGAGGCAGGC
 RC19   TACCTGAAGT CGCCCAAGCC AACC...... GCAA...... GGAGGCAGGC
SBR2016TACCTGAAGT CGCCCAAGCC AACC...... GCAA...... GGAGGCAGGC
 RC7    TACCTGAAGT CGCCCAAGCC AACC...... GCAA...... GGAGGCAGGC
 RC14   TACCTGAAGT CGCCCAAGCC AACC...... GCAA...... GGAGGCAGGC
 RC99   TACCTGAAGT CGCCCAAGCC AACC...... GCAA...... GAAGGCAGGC
 RC11   TACCTGAAGT CGCCCAAGCC AACC...... GCAA...... GGAGGCAGGC
 RC73   TACCTGAAGT CGCCCAAGCC AACC...... GCAA...... GGAGGCAGGC
 RC90   TACCTGAAGT CGCCCAAGCC AACC...... GCAA...... GGAGGCANGC
```

Fig. 8 (continued)

```
[  1651                                                   1700 ]
SBR1024GCCCACGGTA TGACCGATGA TTGGG----- ---------- ----------
SBR1015---------- ---------- ---------- ---------- ----------
 GC86  GCCCACGGTA TGACCGATGA TTGGGGTGAA GTCGTAACAA GGTAACCGTA
SBR2046GCCCACGGTA TGACCGATGA TTGGGG---- ---------- ----------
 RC25  GCCCACGGTA TGGCCCGTGA TTGGGGTGAA GTCGTAACAA GGTAACCGTA
 RC19  GCCCACGGTA TGGCCGGTGA TTGGGGTGAA GTCCTAACA- ----------
SBR2016GCCCACGGTA TGGC------ ---------- ---------- ----------
 RC7   GCCCACGGTA TGGCCG---- ---------- ---------- ----------
 RC14  GCCCACGGTA TGGCCGGTGA T--------- ---------- ----------
 RC99  GCCCACGGTA TGGCCGGTGA ---------- ---------- ----------
 RC11  GCCCACGGTA TGGCCGGTGA TGGGG----- ---------- ----------
 RC73  GCCCACGGTA TGGCCGGTGA TGGGG----- ---------- ----------
 RC90  GCCCACGGTA TGGCCGGTGA TG...----- ---------- ----------

[  1701                                                   1750 ]
SBR1024---------- ---------- ---------- ---------- ----------
SBR1015---------- ---------- ---------- ---------- ----------
 GC86  ATC------- ---------- ---------- ---------- ----------
SBR2046---------- ---------- ---------- ---------- ----------
 RC25  AA-------- ---------- ---------- ---------- ----------
 RC19  ---------- ---------- ---------- ---------- ----------
SBR2016---------- ---------- ---------- ---------- ----------
 RC7   ---------- ---------- ---------- ---------- ----------
 RC14  ---------- ---------- ---------- ---------- ----------
 RC99  ---------- ---------- ---------- ---------- ----------
 RC11  ---------- ---------- ---------- ---------- ----------
 RC73  ---------- ---------- ---------- ---------- ----------
 RC90  ---------- ---------- ---------- ---------- ----------
 ;
```

Fig. 8 (continued)

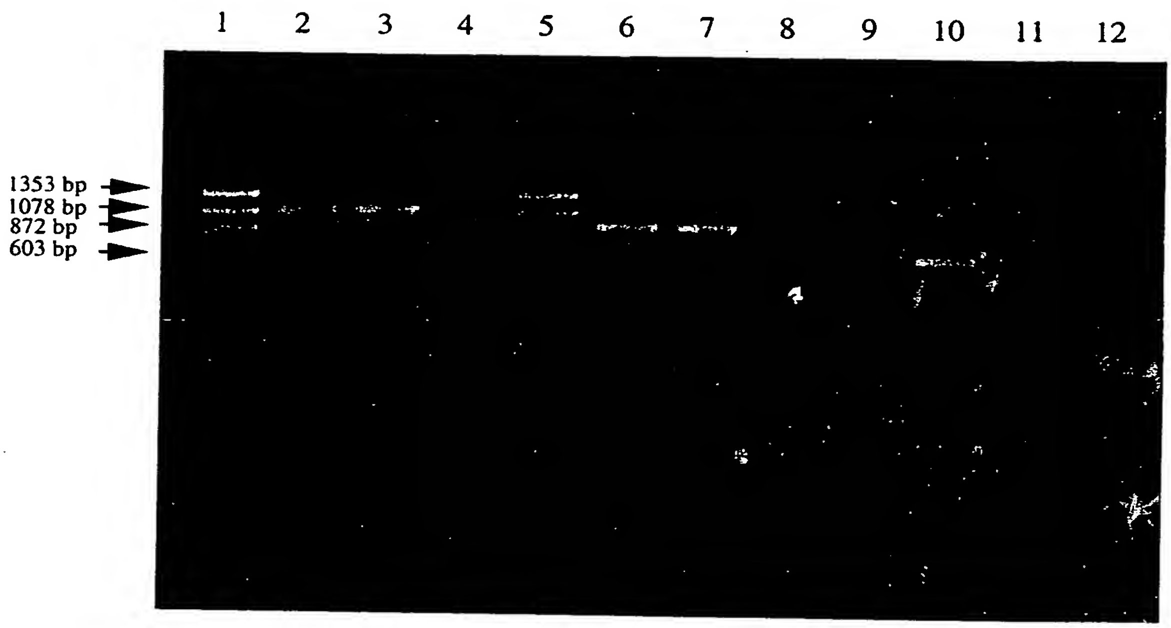


Fig. 9